



**Macromolecular characterization of the
lichen *Peltigera membranacea* using
nucleic acid sequencing
&
Identification of a new lichen species
*Peltigera islandica***

Sheeba Santhini Basil



**Faculty of Life and Environmental Sciences
University of Iceland
December 2018**

**Macromolecular characterization of the lichen
Peltigera membranacea using
nucleic acid sequencing
&
Identification of a new lichen species
*Peltigera islandica***

Sheeba Santhini Basil

**Dissertation submitted in partial fulfillment of a
Philosophiae Doctor Degree in Biology**

Advisor

Prof. Dr. Ólafur S. Andrésson

Ph. D. committee

Prof. Dr. Ólafur S. Andrésson

Dr. Vivian P. W. Miao

Prof. Dr. Zophonías O. Jónsson

Dr. Ólafur Héðinn Friðjónsson

Opponents

Prof. Dr. Imke Schmitt

Prof. Dr. Kristinn P. Magnússon

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, December 2018

Macromolecular characterization of the lichen *Peltigera membranacea* using nucleic acid sequencing & identification of a new lichen species *Peltigera islandica*.

Peltigera membranacea transcriptome (short title).

Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Biology

Copyright © Sheeba Santhini Basil, 2018

All rights reserved

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugata 7
101, Reykjavik
Iceland

Telephone: +(354)- 525 4000

Bibliographic information:

Sheeba Santhini Basil, 2018. Macromolecular characterization of the lichen *Peltigera membranacea* using nucleic acid sequencing & identification of a new lichen species *Peltigera islandica*, Ph. D. dissertation, Faculty of Life and Environmental Sciences, University of Iceland.

ISBN 978-9935-9383-8-1

Printing: Háskólaprent ehf.

Reykjavik, Iceland, December 2018

ABSTRACT

Lichens are a classic example of mutualism, where both partners benefit from each other. The mycobiont serves the photobiont by providing shelter, nutrients, and moisture from the environment and receives carbohydrates and sometimes fixed nitrogen from the photosynthetically active photobiont. Molecular analyses can provide insights into the mechanisms of mutual symbiosis in the cyanolichen *Peltigera membranacea*, but since the lichen is recalcitrant to culture, the global macromolecular components such as DNA, RNA, and proteins must be isolated from field samples. An annotated draft genome of the *P. membranacea* mycobiont is presented, accompanied by the analysis of tissue-specific transcriptomes and methylomes. The assembly consists of 3,033 scaffolds with a total genome size of ~43 Mbp, N50 of 34.52 kbp, and a G+C content of 44.4%. Transcriptomes were generated from about 43 million Illumina sequence reads obtained from three lichen tissues, apothecia, rhizines, and thallus, collected at three-time points. Using the Trinity assembler, initial *de novo* transcriptomes with a total of 110,092 transcript sequences were obtained with lengths ranging from 200 bases to 18,258 bases. Transcriptome data was incorporated into the AUGUSTUS software pipeline and was used to generate a draft annotation (P.mem v1.1) with 16,400 protein-coding models with gene descriptions, gene ontology terms, InterPro identifiers and EC (enzyme code) numbers. About 2,505 proteins were assigned to 112 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Eukaryotic core proteins were identified using BUSCO (Benchmarking Universal Single-Copy Orthologs), showing that the *P. membranacea* genome assembly (P.mem v1.1) represents 96.9% of the BUSCO genes. Furthermore, targeted protein-based searches identified key genes involved in symbiosis, DNA methylation, RNA-mediated silencing, the glyoxylate cycle, melanin pathways, sexual reproduction and nutrient assimilation, leading to a model for the flow of basic nutrients (carbon and nitrogen). The targeted protein searches establish that a homothallic (selfing) sexual cycle can occur in *P. membranacea*, but also shows evidence of a parasexual cycle. Two galectin-like genes, *lec-1* and *lec-2* were characterized and found to be differentially expressed. *Lec-1* expression is enhanced by the presence

of photobionts whereas *lec-2* is highly polymorphic, suggesting strong positive selection, but apparently not driven by variation in photobiont partners. A high level of cytosine methylation (~13.4%,) was found in the genome with an average 2.7% and 8.2% present in exons and introns, respectively. Although all basic genes related to genome defense in ascomycetes were present, a key gene needed for repeat induced point mutation, *rid*, was absent. As the first comprehensive tissue-specific transcriptome and methylome analysis of a cyanolichen, this study lays a foundation for future functional genomics research on these widespread symbiotic organisms. Additionally, a lichen collected during field sampling was characterized as a new cyanolichen species and named as *Peltigera islandica*.

ÚTDRÁTTUR

Fléttur eru vel þekkt dæmi um samlífi sem báðir aðilar hafa gagn af. Blábakteríur eða grænpörungar sjá um ljóstillífun en sveppurinn er af oftast af fylkingu asksveppa. Sveppirnir veita blábakteríunum skjól, næringu og raka, en fá á móti nýtanlegt kolvetni og nitur. Sameindagreining getur veitt innsýn í eðli samlífis hjá fléttum svo sem himnuskóf (*Peltigera membranacea*), en þar sem ekki er gerlegt að rækta fléttuna þarf að einangra stórsameindir svo sem kjarnsýrur og prótein úr sýnum fengnum beint úr náttúrunni. Hér er gerð grein fyrir erfðamengi svepphluta himnuskófar, ásamt greiningu á umritamengjum og DNA metýleringu. Erfðamengjabútarnir 3.033 eru samtals um 43 miljónir basapara með 44,4% G+C hlutfall. Sett voru saman umritasöfn úr alls 43 miljónum raða úr Illumina raðgreiningum. Söfnin voru fengin ur þrenns konar vefjum; þali, rætlingum og askhirslum, og frá þremur tímaskeiðum. Trinity samröðunarforritið gaf samtals 110.092 umrit með lengdir frá 200 til 18.258 basa. Upplýsingar um umrit voru síðan nýtt í forritinu AUGUSTUS til að útbúa glósað erfðamengi (Pmem v1.1) með 16.400 próteintáknandi genum ásamt upplýsingum um varðveitt amínósýrustef og líklega ensímvirkni (EC númer). Alls 2.505 afleidd prótein voru flokkuð á 112 efnaskiptaferla (KEGG – Kyoto Encyclopedia of Genes and Genomes). Lykilprótein heilkjörnunga voru greind með BUSCO (Benchmarking Universal Single-Copy Orthologs). Erfðamengi himnuskófar (Pmem v1.1) inniheldur 96,9% af BUSCO genunum. Þá leiddi leit með próteinum úr skyldum lífverum í ljós lykilgen í næringarnámi, samlífi, DNA metýleringu, RNA miðlaðri þöggun, glyoxalat ferlinu, melanín myndun og í ferlum kynæxlunar. Út frá þessu var lagt fram líkan fyrir flæði næringarefna (kolefni og nitur) í himnuskóf. Einnig benda fengnar upplýsingar til þess að himnuskóf sé einsleit (homothallic) og geti einnig fjölgað sér með hliðæxlun (parasexual). Tvö gen, *lec-1* og *lec-2*, sem tákna prótein sem líkjast galektínunum voru skoðuð og sýnt að þau hafa breytilega tjáningu. Blábakteríur í þali virðast hafa áhrif á tjáningu *lec-1*, en afurðir *lec-2* hafa mjög breytilega amínósýruröð, sem bendir til sterkra valkrafta, sem samsvara þó ekki mismunandi gerð blábaktería í sambýlinu. Mikil cytósín metýlering (~13.4%) er í

erfðamenginu, en metýling í útröðum og innröðum gena er um 2,7% og 8,2%. Í erfðamengið vantar genið *rid*, sem er lykilgen fyrir varnir gegn stökklum og endurteknum röðum, og engin merki fundust um basabreytingar af þess völdum. Þetta er fyrsta víðtæka rannsóknin á stórsameindum í fléttum með blábakteríum (cyanolichen) og leggur grunn að frekari sameindarannsóknum á þessu sviði. Í viðbót við ofangreindar rannsóknir fannst flétta sem við nánari skoðun reyndist vera ný tegund og var hún nefnd foldarskóf (*Peltigera islandica*).

Dedicated to

My Family

*My beloved parents, dear husband, brother and my little princess -
my blessing, my true joy ..*

TABLE OF CONTENTS

ABSTRACT-----	v
ÚTDRÁTTUR -----	vii
TABLE OF CONTENTS-----	xi
LIST OF FIGURES-----	xv
LIST OF TABLES -----	xix
LIST OF ORIGINAL PUBLICATIONS -----	xxii
AUTHOR CONTRIBUTIONS -----	xxiii
GLOSSARY & ABBREVIATIONS-----	xxv
ACKNOWLEDGEMENTS-----	xxix
CHAPTER 1-----	1
1 Introduction -----	3
1.1 Lichens-----	3
1.2 <i>Peltigera membranacea</i> -----	4
1.3 Symbiosis-related genes in lichens -----	5
1.3.1 Lectins -----	6
1.3.2 Hydrophobins-----	6
1.4 Gene silencing in fungi -----	8
1.4.1 Transcriptional gene silencing -----	8
1.4.2 Post-transcriptional gene silencing -----	10
1.5 Technological background -----	13
1.5.1 Transcriptome–RNA sequencing (RNA–Seq) -----	13
1.5.2 Methylome– Bisulfite sequencing (BS–Seq)-----	15
1.6 Omics in lichens -----	17
1.6.1 Metagenomic snapshot of the <i>Peltigera membranacea</i> associated microbiome-----	19
1.7 Objectives-----	25
1.8 Overview of chapters-----	26
CHAPTER 2-----	29

2	THE TISSUE-SPECIFIC TRANSCRIPTOMIC LANDSCAPE OF THE LICHEN	
	<i>PELTIGERA MEMBRANACEA</i>	31
2.1	Introduction	31
2.2	Materials and Methods	34
2.2.1	Lichen Material and RNA extraction	34
2.2.2	RNA-Seq library preparation	35
2.2.3	RNA-Seq reads processing	36
2.2.4	Generation of gene models, gene prediction and annotation, and metabolic pathway identification	36
2.2.5	Identification and analysis of genes	40
2.2.6	Expression analysis	40
2.3	Results	42
2.3.1	Transcriptome sequencing and assembly	42
2.3.2	BUSCO analysis	47
2.3.3	Data quality control and gene prediction	50
2.3.4	Functional annotation and metabolic pathways	59
2.3.5	Identification of Differentially Expressed Genes (DEGs)	68
2.3.6	Analysis of specific genes and pathways	79
2.3.7	Transporters in <i>P. membranacea</i>	82
2.3.8	Melanin pathway	83
2.3.9	Glyoxylate pathway	84
2.3.10	Genes implicated in symbiosis	87
2.3.11	Identification of Mating-type and heterokaryon incompatibility genes	88
2.4	Discussion	91
2.4.1	RNA sequencing	91
2.4.2	Estimation of the genome assembly completeness using BUSCO	92
2.4.3	Genome annotation and validation	92
2.4.4	Gene ontology (GO) and metabolic pathway analysis	93
2.4.5	Expression and metabolic pathway analysis	94
2.4.6	Nutrient assimilation	95
2.4.7	Glyoxylate pathway	99
2.4.8	Melanin pathway	100
2.4.9	Symbiotic genes	100
2.4.10	Mating-type and heterokaryon incompatibility genes	103
2.4.11	Caveats to the study	104
2.5	Conclusion	105

CHAPTER 3	107
DNA CYTOSINE METHYLATION AND GENOME DEFENSE IN THE LICHEN	
<i>PELTIGERA MEMBRANACEA</i>	107
3 DNA CYTOSINE METHYLATION AND GENOME DEFENSE IN THE	
LICHEN <i>PELTIGERA MEMBRANACEA</i>	109
3.1 Introduction	109
3.2 Materials and Methods	115
3.2.1 Lichen Material, DNA and RNA isolation	115
3.2.2 Background of <i>P. membranacea</i> genome	116
3.2.3 Bisulfite treated whole genome DNA sequencing (WGBS)	116
3.2.4 BS-Seq data processing	116
3.2.5 Transcriptome sequencing and analysis	120
3.2.6 Southern blotting	121
3.2.7 Identification and manual annotation of genes	121
3.2.8 RIP analysis of <i>P. membranacea</i> repeat families	121
3.3 Results	122
3.3.1 DNA methylome of <i>P. membranacea</i> tissues	122
3.3.2 Confirmation of DNA cytosine methylation by Southern blot	
analysis	131
3.3.3 Absence of RID and repeat-induced point mutations	134
3.3.4 DNA methylation and expression of transposable elements	
(TE), an inverse relationship	135
3.3.5 RNA silencing genes	136
3.3.6 Identification of genes encoding DNA methyltransferases	136
3.4 Discussion	139
3.5 Conclusion	143
CHAPTER 4	145
GALECTIN-LIKE GENES: IDENTIFICATION AND EXPRESSION OF <i>LEC-1</i> AND	
<i>LEC-2</i> IN <i>PELTIGERA MEMBRANACEA</i>	145
(PAPER 1 AND PAPER 2)	145
PAPER 1	147
EXPRESSION OF <i>LEC-1</i>, A MYCOBIONT GENE ENCODING A GALECTIN-LIKE	
PROTEIN IN THE LICHEN <i>PELTIGERA MEMBRANACEA</i>	147
PAPER 2	159

LEC-2, A HIGHLY VARIABLE LECTIN IN THE LICHEN <i>PELTIGERA</i>	
<i>MEMBRANACEA</i>-----	159
CHAPTER 5-----	169
<i>PELTIGERA ISLANDICA</i>, A NEW CYANOLICHEN SPECIES IN SECTION	
<i>PELTIGERA</i> ('<i>P. CANINA</i> GROUP')-----	169
PAPER 3-----	171
<i>PELTIGERA ISLANDICA</i>, A NEW CYANOLICHEN SPECIES IN SECTION	
<i>PELTIGERA</i> ('<i>P. CANINA</i> GROUP')-----	171
4 SUMMARY AND OUTLOOK-----	190
5 REFERENCES-----	193
APPENDIXES-----	219
Appendix A: <i>Peltigera membranacea</i> sample collection, DNA isolation	
and genome features.-----	221
Appendix B: Metagenomic analysis.-----	224
Appendix C: BUSCO analysis.-----	241
Appendix D: CEGMA Analysis.-----	243
Appendix E: Comparison of the <i>de novo</i> transcriptome assembly and	
reference genome annotation.-----	246
Appendix F: Alignment of predicted and curated proteins.-----	254
Appendix G: KEGG analysis.-----	270
Appendix H: List of DEGs from all the tissue comparisons.-----	295
Appendix I: Southern blotting protocol.-----	397
Appendix J: Analyses of repeat elements, RIP indices and <i>rid</i> gene	
identification in <i>P. membranacea</i> genome.-----	399
Appendix K: BS-Seq analysis.-----	403
Appendix L: KEGG maps.-----	442
Appendix M: blast2go2gff.pl-----	448

LIST OF FIGURES

Figure 1.1 Morphology of the foliose lichen <i>Peltigera membranacea</i>	5
Figure 1.2 Distribution (bold lines) of hydrophobins in the lichen thallus.	7
Figure 1.3 Possible mechanisms for C to T mutation by RIP.....	9
Figure 1.4 Gene silencing pathways in <i>Neurospora</i>	12
Figure 1.5 RNA–Seq.	14
Figure 1.6 Schematic of Bisulfite conversion of Cytosine to Uracil.....	15
Figure 1.7 BS–Seq.	16
Figure 1.8 Mapping of reads to <i>P. membranacea</i> ITS1–5.8S–ITS2 region.	20
Figure 1.9 Mapping of reads to <i>Cyphobasidium usneicola</i> ITS1–5.8S–ITS2 regions.	22
Figure 2.1 Denaturing RNA gel electrophoresis of total RNA.	35
Figure 2.2 Genome annotation and RNA–Seq analysis workflow.	39
Figure 2.3 Length distribution of <i>de novo</i> assembled transcripts.....	46
Figure 2.4 BUSCO analysis.	49
Figure 2.5 Number of missing and unmapped <i>de novo</i> transcripts from P.mem v1.1 genome.	51
Figure 2.6 Alignment of predicted and curated proteins.	57
Figure 2.7 Mapping of PKS genes to P.mem v1.1 genome.	58
Figure 2.8 Summary of Blast2GO analysis.	59
Figure 2.9 Distribution of P.mem v1.1 sequences by Gene Ontology (GO) at hierarchy level 2.	60
Figure 2.10 Two views of KEGG maps of metabolic pathways present in <i>Peltigera membranacea</i>	67
Figure 2.11 Cumulative distribution (CD) plots of thallus whole technical replicates collected in summer 2012.....	68
Figure 2.12 Datastore tree of the different datasets.....	71
Figure 2.13 Heatmap from hierarchical clustering of Apothecia (autumn 2009) vs. rhizine (autumn 2010 and summer 2012) vs. thallus body (summer 2012).	72
Figure 2.14 BLASTP analysis of DEGs.	74
Figure 2.15 Blast2GO analysis of DEGs.	75

Figure 2.16 Bar Chart of GO annotation of DEGs annotated to the different GO terms at hierarchy level 2 for all the three GO categories.	77
Figure 2.17 Enzyme classification for DEGs.	78
Figure 2.18 KEGG map of the Glyoxylate (and TCA) cycle identified in <i>Peltigera membranacea</i>	86
Figure 2.19 Proposed model for nutrient flow (C and N) in <i>P. membranacea</i>	98
Figure 3.1 Heterochromatin formation and DNA methylation in <i>N. crassa</i>	111
Figure 3.2 sRNA biogenesis pathways and RNAi functions in <i>Neurospora</i>	114
Figure 3.3 BS-Seq workflow, from generation of reads to mapping of the reads.	119
Figure 3.4 DNA cytosine methylation levels in <i>P. membranacea</i>	124
Figure 3.5 Methylation and expression levels in exons.	126
Figure 3.6 Screenshot showing exon methylation.	127
Figure 3.7 Methylation in introns.	129
Figure 3.8 Screenshot showing intron methylation.	130
Figure 3.9 Southern blotting.	132
Figure 3.10 Validation of DNA methylation using Bisulfite sequencing.	133
Figure 3.11 Sequence alignment of DIM-2 of <i>P. membranacea</i>	138
Figure S1: Relative abundance of OTUs of bacterial communities present in the apothecia of <i>Peltigera membranacea</i>	228
Figure S2: Relative abundance of OTUs of bacterial communities present in the thallus of <i>Peltigera membranacea</i>	230
Figure S3: Multiple alignment of 18S rDNA subunits.	232
Figure S4: Mapping of 454 apothecial reads to 18S rDNA subunit.	236
Figure S5: RNA analysis of <i>P. membranacea</i> RNA on the Agilent Bioanalyzer 2100.	237
Figure S6: Heat map showing 3' bias.	240
Figure S7: Sequence alignment of Glucose-6-phosphate isomerase.	254
Figure S8: Sequence alignment of aldehyde dehydrogenase.	255
Figure S9: Sequence alignment of high affinity ammonium transporter.	255
Figure S10: Sequence alignment of galectin like protein (<i>lec-1</i>).	256
Figure S11: Sequence alignment of galectin like protein (<i>lec-2</i>).	257
Figure S12: Sequence alignment of RNA polymerase II large subunit, partial.	259
Figure S13: Sequence alignment of Elongation factor-1 alpha.	260

Figure S 14: Sequence alignment of peptidase A22B	261
Figure S15: Sequence alignment of mating-type switching protein swi10, partial.....	262
Figure S16: Sequence alignment of TBP (TATA binding protein) associated factor.	264
Figure S17: Sequence alignment of mitochondrial carrier protein.....	265
Figure S18: Sequence alignment of DNA directed RNA polymerase II. .	267
Figure S19: Sequence alignment of heat shock protein (<i>hsp98</i>).	268
Figure S20: Sequence alignment of beta-tubulin (<i>tub-2</i>).....	269
Figure S21: Heatmaps for hierarchical clustering of DEGs for the different tissue comparisons.	289
Figure S22: PCA plots of DEGs for the different tissue comparisons.	294
Figure S23: Exon methylation level for each sequence contexts.....	413
Figure S24: KEGG map of Biosynthesis of secondary metabolites identified in <i>Peltigera membranacea</i>	442
Figure S25: KEGG map of Biosynthesis of amino acids identified in <i>Peltigera membranacea</i>	443
Figure S26: KEGG map of Purine metabolism identified in <i>Peltigera membranacea</i>	444
Figure S27: KEGG map of Carbon metabolism identified in <i>Peltigera membranacea</i>	445
Figure S28: KEGG map of Oxidative phosphorylation identified in <i>Peltigera membranacea</i>	446
Figure S29: Streptomycin biosynthesis pathway in <i>P. membranacea</i> . ..	447

LIST OF TABLES

Table 1.1 Omics studies on lichens.	18
Table 1.2 Summary of reads mapped.	21
Table 1.3: Summary of BLASTN hits of 454 reads.	21
Table 1.4: Taxonomic assignments of 454 reads using the SILVA database (v132).	24
Table 2.1 Pairwise comparisons of the three tissues collected at different time points.	42
Table 2.2 Summary of RNA–Seq experimental results.	44
Table 2.3 List of curated proteins used in the gene prediction validation study.	53
Table 2.4 List of the metabolic pathways identified in <i>P. membranacea</i>	62
Table 2.5 Summary of DEGs.	69
Table 2.6 List of genes and enzymes encoded in <i>P. membranacea</i> involved in nitrogen and carbon metabolism.	81
Table 2.7 List of genes involved in the melanin biosynthesis pathway. ..	84
Table 2.8 List of genes and enzymes encoded in <i>P. membranacea</i> involved in glyoxylate cycle.	85
Table 2.9 List of symbiosis-related genes in <i>P. membranacea</i>	87
Table 2.10 List of mating, heterokaryon incompatibility, vegetative compatibility, and sexuality-related genes of <i>P. membranacea</i> genes and their homologs in <i>Aspergillus nidulans</i> , <i>Neurospora crassa</i> and <i>Podospora anserina</i>	89
Table 3.1 Summary of BS–Seq sequencing.	122
Table 3.2 RIP indices for the transposable element families in <i>P. membranacea</i> and several other ascomycetes.	134
Table 3.3 DNA methyltransferases and genes involved in RNA interference in <i>P. membranacea</i>	137
Table S1: <i>Peltigera membranacea</i> genome statistics.	223
Table S2: Summary of <i>P. membranacea</i> 454 reads.	225
Table S3: Summary of classified reads.	226
Table S4: Summary of transcriptome assemblies.	240

Table S5: BUSCO analysis.....	241
Table S6: Analysis of missing BUSCO proteins from <i>P. membranacea</i> genome (P.mem v1.1).	242
Table S7: CEGMA report on <i>P. membranacea</i> genome assembly P.mem v1.1).	244
Table S8: CEGMA report on <i>P. membranacea</i> Trinity transcriptome assembly.	244
Table S9: CEGMA report on <i>Aspergillus nidulans</i> genome.	245
Table S10: Presence/absence of missing CEGMA proteins from <i>P. membranacea</i> genome (P.mem v1.1), <i>P. membranacea</i> Trinity transcriptome and <i>A. nidulans</i> genome.	245
Table S11: List of missing, and unmapped <i>de novo</i> Trinity transcripts in P.mem v1.1 genome along with BLAST hits.....	246
Table S12: BLAST results for the unmapped Rhizine (au-2009) Trinity <i>de novo</i> transcripts	246
Table S13: BLAST results for the unmapped apothecia (au-2009) Trinity <i>de novo</i> transcripts	249
Table S14: BLAST results for the unmapped Thallus whole (au-2009) Trinity <i>de novo</i> transcripts.	252
Table S15: BLAST results for the unmapped Thallus whole (su-2011) Trinity <i>de novo</i> transcripts	253
Table S16: List of metabolic pathways removed from <i>P. membranacea</i>	270
Table S17: The KEGG pathway table of <i>P. membranacea</i>	271
Table S18: List of 768 DEGs between TW–B vs. TW–G vs. TW–W vs. TW–R and TW–Y (su 12). The background variation genes i.e., genes with >2X fold change are highlighted in red.	295
Table S19: List of 413 DEGs between A (au–09) vs. TB (su–12) vs. R (su–12), and the 74-background variation DEGs are highlighted in red.....	331
Table S20: List of 265 DEGs between A (au–09) vs. TW (au–09), and the 56-background variation DEGs are highlighted in red.....	345
Table S21: List of 219 DEGs between A (au–09) vs. R (au–10).	353
Table S22: List of 183 DEGs between R (au–10) vs. TW (au–10), and the 65-background variation DEGs are highlighted in red.	360
Table S23: List of 84 DEGs between A (au–09) vs. TB (su–12), and the 11-background variation DEGs are highlighted in red.	366
Table S24: List of 294 DEGs between R (su–12) vs. TB (su–12), and the 63-background variation DEGs are highlighted in red.	369

Table S25: List of 317 DEGs between TW (au–09) vs. TW (au–10) vs. TW (Su–11) vs. TW (su–12), and the 105-background variation DEGs are highlighted in red.....	377
Table S26: List of 76 DEGs between R (au–10) vs. R (su–12).....	390
Table S27: List of 34 DEGs between TB (su–12) vs. TW–G (su–12), and the 7-background variation DEGs are highlighted in red.....	393
Table S28: List of dioxygenases and lipases genes in <i>P. membranacea</i>	395
Table S29: Repeat elements in <i>P. membranacea</i> genome.	399
Table S30: RIP analysis.....	401
Table S31: <i>rid</i> (RIP defective) homolog in <i>P. membranacea</i> and other fungi.....	402
Table S32: Summary of BS–Seq sequencing.	403
Table S33: Overview of BS reads mapped.	404
Table S34: DNA cytosine methylation levels in rhizines (R), thallus body (TB) and thallus whole (TW).	405
Table S35: List of genes and the methylated sequence contexts of differentially methylated regions for exons.	406
Table S36: Methylation and expression levels in exons	409
Table S37: List of scaffolds and the methylated sequence contexts of differentially methylated regions for introns.	414
Table S38: Methylation and expression level in introns.....	423
Table S39: TE with expression and methylation level.	431
Table S40: List of all the transporters in <i>P. membranacea</i> and their expression in A, TB and R tissues.....	432
Table S41: List of PKSs in <i>P. membranacea</i> (P.mem v1.1) genome.	449

LIST OF ORIGINAL PUBLICATIONS

The thesis is comprised of four chapters. Chapter 1 is the general background, chapters 2 and 3 are unpublished manuscripts and chapters 4 and 5 are based on three published papers. The papers and manuscripts are referred in the text by Roman numerals:

- I. Manoharan–Basil SS, Andr sson  S.** The tissue-specific transcriptomic landscape of the lichen *Peltigera membranacea*. *Manuscript. (Chapter 2)*
- II. Manoharan–Basil SS, Miao VPW, Andr sson  S.** DNA cytosine methylation and genome defense in the lichen *Peltigera membranacea*. *Manuscript. (Chapter 3)*
- III. Miao VPW, Manoharan SS, Sn bjarnarson V, Andr sson  S (2012).** Expression of *lec-1*, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea*. *Symbiosis* 57: 23–31. doi: 10.1007/S13199-012-0175-1. **(Chapter 4)**
- IV. Manoharan SS, Miao VPW, Andr sson  S (2012).** LEC-2, a highly variable lectin in the lichen *Peltigera membranacea*. *Symbiosis* 58: 91–98. doi: 10.1007/s13199-012-0206-y. **(Chapter 4)**
- V. Manoharan–Basil SS, Miadlikowska J, Goward T, Andr sson  S and Miao VPW.** *Peltigera islandica*, a new cyanolichen species in section *Peltigera* ('*P. canina* group'). *The Lichenologist* 48(5): 451–467 (2016). doi:10.1017/S0024282916000414. **(Chapter 5)**

All papers are reprinted with permission from the publishers.

AUTHOR CONTRIBUTIONS

- I. SSM-B collected and processed the samples, performed a cross section of *P. membranacea*, isolated DNA and RNA, prepared the libraries for RNA-Seq, performed all the bioinformatic analysis, designed the model for the flow of carbon and nitrogen, and wrote the manuscript. (*Chapter 2*)
- II. SSM-B collected and processed the samples, isolated DNA, prepared the libraries for BS-Seq, performed all the experiments and the bioinformatic analysis, and wrote the manuscript. (*Chapter 3*)
- III. SSM collected and processed the samples, performed the experiments, and took part in the data analysis and writing of the article. (*Chapter 4*)
- IV. SSM initiated and designed the project, collected and processed the samples, performed the experiments, analysed the data and took part in writing the article. (*Chapter 4*)
- V. SSM-B initiated and designed the project, collected and processed the samples, isolated DNA, performed a cross section of *P. islandica*, cultured the cyanobiont, carried out sequencing, took part in data analysis and wrote the initial draft with subsequent input from all the authors. (*Chapter 5*)

GLOSSARY & ABBREVIATIONS

3D	In three dimensions
5mC	5-methyl cytosine
A	Apothecia
aa	Amino acids
AAT	Aspartate aminotransferases
AlaAt	Alanine aminotransferase
allele	Each of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome.
aLRT	Approximate likelihood-ratio test
AMT	Ammonium transporters
BLAST	Basic local alignment search tool
BP	Biological process
BS-Seq	Bisulfite sequencing
C	Cytosine
C3	Calvin-Benson cycle
CC	Cellular component
CD	Cumulative distribution
cDNA	Complementary DNA
CEG	Core eukaryotic genes
CMT	Chromomethylases
CPSI	Carbamoyl phosphate synthetase
CRD	Carbohydrate recognition domain
CTOB	Complementary to the original bottom (OB) strand
CTOT	Complementary to the original top (OT) strand
DEG	Differentially expressed genes
DIM	Defective in DNA methylation
DMM	DNA methylation modulator

DMT	DNA methyltransferase
DNA	Deoxyribonucleic acid
EC	Enzyme code
FKPM	Fragments per kb transcript per million fragments mapped
GDH	Glutamate dehydrogenase
GI	Genbank identification
GO	Gene ontology
GOGAT	Glutamate synthase (Glutamine oxoglutarate aminotransferase)
GPD	Glyceraldehyde-3-phosphate dehydrogenase
GPI	Glycosylphosphatidylinositol
GPT	Glutamate pyruvate transaminase
GS	Glutamine synthetase
H1	Hydrophobin
Heterothallism	Fungi with two compatible mating types are required for a complete sexual cycle
HMG	High mobility group
Homothallism	Fungi can undergo sexual reproduction and form sexual structures and progeny during solo growth
IPS	InterProScan
ITS	Internal transcribed spacer (between rRNA genes)
KEGG	Kyoto encyclopedia of genes and genomes
KOGs	Eukaryotic orthologous groups
LEC	Lectin
<i>Mat</i> locus	Mating-type locus
MEP	Methylammonium permeases
MF	Molecular function
MIP	Methylation induced premeiotically
ML	Maximum likelihood
mRNA	Messenger RNA
MSUD	Meiotic silencing of unpaired DNA
NGS	Next generation sequencing
nt	Nucleotide

OB	Original bottom strand
OT	Original top strand
PE	Paired end
PKS	Polyketide synthase
Pmb-WGS	<i>P. membranacea</i> whole genome sequencing
Pseudohomothallism	Fungi package two opposite mating-type nuclei in the same spore
PTGS	Posttranscriptional gene silencing
RdDM	RNA-dependent DNA methylation
RdRP	RNA dependent RNA polymerase
Retrotransposons	Transposons that have RNA intermediates
RID	RIP defective
RIP	Repeat-induced point mutation
RISC	RNA-induced silencing
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNAi	RNA interference
RPB1	RNA polymerase II subunit 1
RubisCO	Ribulose 1, 5 biphosphate carboxylase/ oxygenase
SAM	S-adenosylmethionine
SGOT	Glutamate oxaloacetate transaminase
TB	Thallus body
TCA	Tricarboxylic acid
TE	Transposable element
TGS	Transcriptional gene silencing
Transposons	Sequences of DNA that can move or transpose themselves to new positions within the genome
TUB-2	Beta tubulin
TW	Thallus whole (thallus with rhizines)

ACKNOWLEDGEMENTS

This research project would not have been possible without the support of many people. I would like to extend my deepest gratitude, especially to the following individuals. First, I thank my supervisor Prof. Dr. Ólafur S. Andr sson whose wisdom has been invaluable to me and provided me with this research opportunity. Deepest gratitude to the members of the supervisory committee Dr. Vivian P. W. Miao, Dr.  lafur H  inn Fri j nsson and Dr. Zophon as O. J nsson. I would also like to give my appreciation to Dr. Daniele Armaleo, Duke University, U.S.A for his always enlightening and encouraging words. I would like to particularly acknowledge the help of late Sigr  ur H.  orbjarnard tti for helping me familiarize with the lab codes. Thank you, Dr. Jolanta Miadlikowska, Duke University, U.S.A and Naturalist Trevor Goward, UBC, Canada for collaborating on the identification of the new lichen, *Peltigera islandica*. I am much obliged to Prof. Dr. Surbhi Malhotra-Kumar, UA, Belgium for allowing me to access CLC Genomics Workbench, Dr. V ronique Nussenblatt, for the stroll around Bist, the Belgian fries and proofreading Chapter 2,  lafur  . Magn sson (deCODE Genetics, Inc.) for help with sequencing, Prof. Dr. Matteo Pellegrini and Dr. Marco Morselli, UCLA, U.S.A for valuable discussion on methylation analysis, and thank you, Dr. Simon Andrews, for troubleshooting with the Seqmonk analysis. This thesis and research would not be realistic without funding, so thanks to The Icelandic Research Fund for funding this project. Finally, deep gratefulness goes to my princess Benita Samara Basil, mom (Sama), dad (Manoharan), hubby (Britto)-for their unconditional love and support throughout my life, brother (Marshall) - for always being willing to babysit Benita    while I work on the thesis, Fr. Dr. Amaldoss and Silvia nanny, parent-in-law (Elizabeth & Xavier) and sisters-in-law for their constant support through prayers.

Last but not least, I thank the Lord Almighty for the wisdom and perseverance that he has bestowed upon me during this research, and throughout my life. You have made my life more bountiful. May your name be honored and glorified.

CHAPTER 1

INTRODUCTION

1 Introduction

The macromolecular contents of cells and organisms form a rich informational foundation for modern biological research in fields as diverse as human medicine, agriculture, evolution, and ecology. Technological advances have now made it possible to obtain data for the genome and transcriptome of non-model organisms. Additional global features such as methylation patterns and other epigenetic or posttranscriptional modifications can also be characterized and added to the global model. Global macromolecular analysis can thus provide insights into the basis of biological phenomena such as mutualistic symbiosis, often epitomized by lichens. A lichen is a mutualistic symbiotic organism, and the lichen symbiosis may be shaped by lichen-derived molecules such as lectins, secondary metabolites and a multitude of other factors that are involved in mycobiont and photobiont interactions. This work uses the model lichen system *Peltigera membranacea* and current tools of molecular biology are used to gain insights into the macromolecular characterization of this lichen.

This chapter briefly describes the background of the research presented in this dissertation along with the objectives and overview of each chapter. It starts with a review of the biology of lichens and the model lichen *Peltigera membranacea*, used in this study. Sections 1.3 and 1.4 provide a synopsis of genes involved in the lichen symbiosis and gene silencing in fungi, respectively. Section 1.5 reviews the next generation sequencing (NGS) technologies used in this study. Section 1.6 discuss prior omics work on lichens, and section 1.6.1 briefly describes the mycobiont genome and the metagenomic composition of *Peltigera membranacea*.

1.1 Lichens

A lichen is a symbiotic association of a fungus (mycobiont,) and one or more photosynthetic partners (photobiont), either a green alga (chlorobiont or phycobiont) or a cyanobacterium (cyanobiont), or both. About 85% of lichen-forming fungi are associated with green algae, about 10% with cyanobacteria and about 5% with both (Nash 2008). The name of a lichen is determined by the species of the mycobiont. A lichen is

referred to as bipartite when the mycobiont has a single photosynthetic partner (chlorobiont or cyanobiont) and tripartite when both types of photobionts are present, where the chlorobiont is the primary photobiont, and the cyanobiont is confined to specialized structures called cephalodia (Nash 2008). Bipartite lichens with chlorobionts are called chlorolichens, and those with cyanobionts are called cyanolichens.

The mycobiont and the photobiont form a composite structure called a thallus (plural thalli), and are generally classified as crustose (a flat crust, commonly seen on walls and other stonework), foliose (leafy), fruticose (branched, shrubby), filamentous (hair-like), leprose (powdery), squamulose (made up of small scales) or gelatinous (generally lichens in which the cyanobacterium produces a polysaccharide that absorbs and retains water) (Büdel and Scheidegger 2008; Sanders 2001).

Lichens grow practically everywhere – on rocks or tree bark, or on poorly developed soils, in deserts and tropical rainforests. Extensive lichen populations also grow at high latitudes in the Subarctic, Arctic, and Antarctic, where lichens are of significant ecological importance and where global warming and climate change pose a fundamental threat. Lichens harbouring cyanobionts fix molecular nitrogen (Nash 2008) and help colonize new lava fields and barren areas that are nutrient deficient (Kristinsson and Heiðmarsson, 2009). Lichens are used as air pollution indicators and are used to map pollution zones around many of the world's major cities and are as such important in monitoring atmospheric pollutants such as sulfur dioxide, ozone, and fluorides (Richardson 1992).

1.2 *Peltigera membranacea*

Peltigera membranacea, the model organism of this study, is a foliose cyanolichen with two major symbionts: the mycobiont, a heterotrophic ascomycete, and the photobiont, a *Nostoc* sp. cyanobacterium. *P. membranacea* has been chosen as a model organism for this study because it is bipartite, is widely distributed, is nitrogen fixing and is representative of lichens common at northern latitudes. In these lichens, the mycobionts associate with photobionts to form a layered symbiotic vegetative structure called the thallus. Within a thallus, the *Nostoc* cells are found in a layer that is otherwise comprised of densely aggregated fungal hyphae. The latter form an upper cortex of more closely packed cells, and a medullary layer of loosely woven hyphae that give rise to aposymbiotic vegetative structures, rhizines, which are hair-like growths that anchor the thallus to its substrates. The mycobiont produces brown or

blackish fruiting bodies, apothecia, aposymbiotic sexual reproductive structures that arise at the margins of some lobes of the thallus (Figure 1.1). In this study the thallus without rhizines is referred to as the thallus body.

Morphology of *Peltigera membranacea*

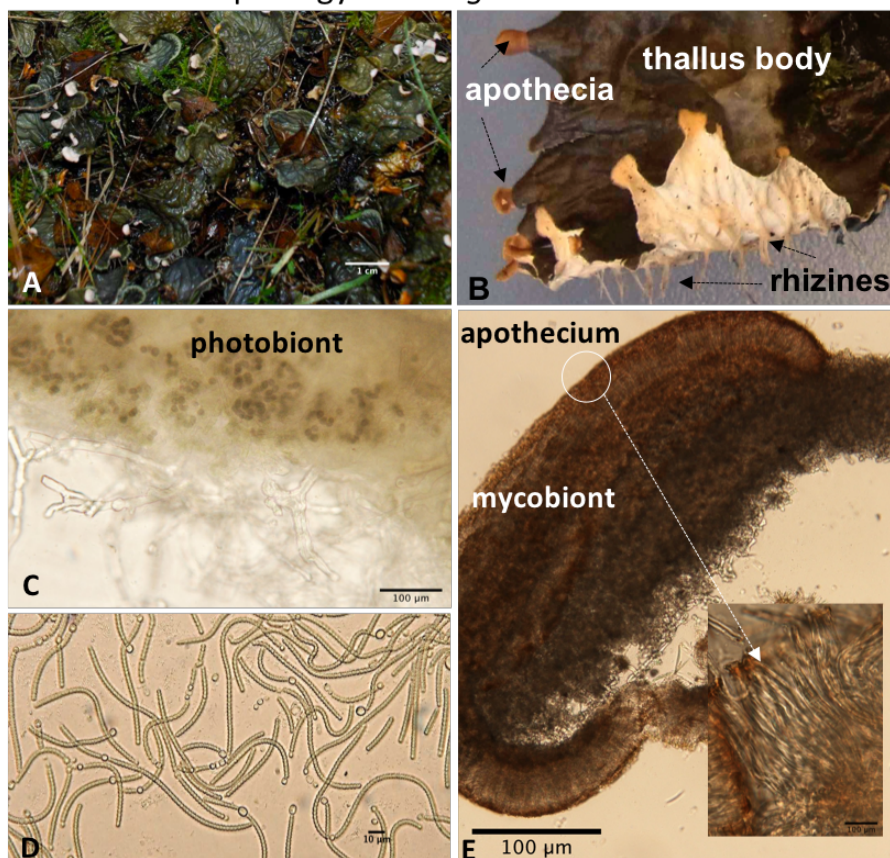


Figure 1.1 Morphology of the foliose lichen *Peltigera membranacea*. (A) Lichen *in situ* (~3.5 cm across) in a moist habitat. (B) Apothecia, thallus body and rhizines (4–8 mm) (6.5 cm across). (C) Cross section of a thallus body (~0.4–0.5 mm thick) showing the stratified arrangement of symbionts. (D) Micrograph showing chains of the cultured photobiont (*Nostoc* sp.). (E) Cross section of apothecia with inset showing asci and supporting cells.

1.3 Symbiosis-related genes in lichens

Symbiosis by definition refers to the long-term partnership established between two or more species. The partnership can be mutualistic,

parasitic, or commensal. When organisms benefit from each other, it is called mutualism. In parasitism, the relationship benefits one organism and lowers the fitness of their host whereas, in commensalism, the host is neither harmed nor helped and is in a neutral relationship (Piel 2010). Lichens exhibit a mutualistic symbiosis. Lichen symbiosis is often initiated and maintained by signaling and direct interaction – the genes involved are often referred to as 'symbiotic genes', where the best-studied examples are lectins and hydrophobins.

1.3.1 Lectins

Lectins are a large group of proteins of diverse structure that have the common property of binding to carbohydrates such as those found on the surface of microorganisms. Many have a well-known role in cell-cell recognition via surface molecule interactions (Sharon 2007): e.g. lectins from legumes (phytoagglutinins) are involved in determining the compatibility of nodulation interactions with particular strains of rhizobia (Mathis et al. 2005); in animals, other lectins (galectins) mediate immune responses (Vasta 2009). In mammals, the galectins form a well-delimited family by their shared affinity for β -galactosides and the presence of conserved residues in the carbohydrate recognition domain (Cooper 2002; Leffler et al. 2004; Cummings and Liu 2009). Most lectins known from fungi have been reported from basidiomycete species, often in the context of potential therapeutic or other practical applications (Lyimo et al. 2011; Singh et al. 2010, 2011), but studies in other taxa have provided additional perspectives (Balogh et al. 2003; Nowrousian and Cebula 2005). In the lichen-forming ascomycetes, the participation of lectins as potential mediators of symbiosis has been of long-standing interest among lichenologists (Elifio et al. 2000; Fontaniella et al. 2004; Galun et al. 1976; Bubrick et al. 1981; Howe and Barrett 1970; Lockhart et al. 1978; Molina et al. 1998; Molina and Vincente 1995; Sacristan et al. 2006). Given their roles in surface recognition and as determinants of specificity in other biological systems, it seems likely that lectins are involved at one or more stages in the interaction between symbionts in lichens. The present study describes the identification of two lectin-like genes, *lec-1* (III) and *lec-2* (IV) from *P. membranacea* based on mycobiont genome assemblies and RNA-Seq.

1.3.2 Hydrophobins

Exchange of gas, water and nutrients is vital to the functioning of the lichen symbiosis. Lichens have a multilayered structure consisting of an outer gelatinous cortex, an upper photobiont layer where the fungal

hyphae entangle the photobiont partner, and an inner gas-filled medulla (Figure 1.2) (Dyer 2002). The inner thallus contains compounds that resist water entry, thereby preventing waterlogging of the lichen and maintaining the thallus 'inner atmosphere' (Lange et al. 1997). Hydrophobins prevent waterlogging of key airspaces within lichen thalli, facilitating the exchange of water and nutrients between the symbiotic partners and thereby isolating the photobiont making them reliant on the mycobiont for water and nutrients (Figure 1.2) (Honegger 1996; Honegger 1991). Hydrophobins are small secreted proteins (100-150 aa) that are hydrophobic and can form an amphipathic film at hydrophilic-hydrophobic interfaces (Wessels 2000; Wessels 1997). The hydrophobin *xph1* gene from *X. parietina* was expressed when the mycobiont and the photobiont grew together, i.e., during symbiosis. Higher expression was seen in the medullary hyphae of the vegetative thallus and the thalline margin of apothecia; these are in close contact with the photobiont layer and prevent the thalline interior from becoming waterlogged, which would impede gas exchange of the photobiont. No signal was detected in either the hydrophilic cortex, the conidiomata, the hymenial and subhymenial layers of apothecia, or in the aposymbiotically cultured mycobiont (Scherrer et al. 2000). The study presented here describes the identification and expression of hydrophobins in *P. membranacea* (I).

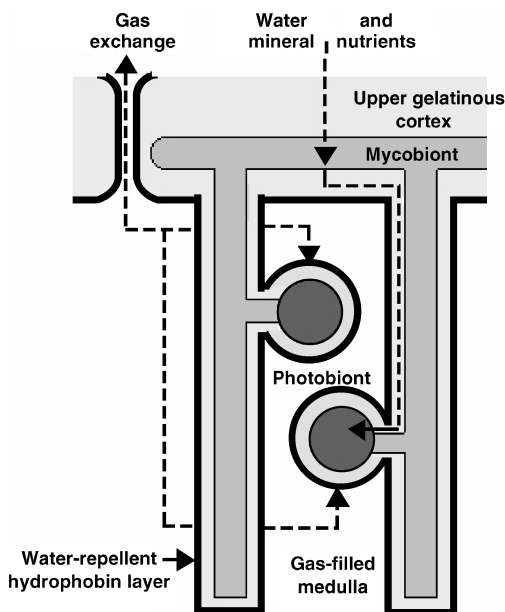


Figure 1.2 Distribution (bold lines) of hydrophobins in the lichen thallus. (Reproduced from Dyer (2002).

1.4 Gene silencing in fungi

Gene silencing can suppress or interrupt the expression of genes and can affect introduced foreign nucleic acids by inducing sequence-specific gene silencing and/or can act on endogenous duplicated sequences (Catalanotto et al. 2000; Cogoni 2001), thereby protecting the genome against disruptive transposons (Cogoni and Macino, 1999). Homology-dependent gene silencing (HDGS) is a ubiquitous phenomenon found among fungi, plants, and animals (Cogoni 2001; Matzke and Antonius 1995). In HDGS, multiple copies of the same or homologous sequences that are introduced into the genome are inactivated/silenced. HDGS can be divided into two categories, transcriptional gene silencing (TGS), when no transcript of the targeted gene is produced, and posttranscriptional gene silencing (PTGS) (Jauvion et al. 2012), where the transcripts are specifically degraded before translation can occur, thereby preventing the accumulation and translation of the transcripts albeit continuing transcription. Both PTGS and TGS mechanisms exist in fungi.

1.4.1 Transcriptional gene silencing

Two TGS mechanisms, Repeat Induced Point mutation (RIP) in *Neurospora crassa* (Selker et al., 1987) and Methylation - Induced Premeiotically (MIP) in *Ascobolus immersus* (Goyon and Faugeron, 1989) have been reported so far in filamentous Ascomycetes (Pezizomycotina).

1.4.1.1 RIP

RIP, originally an acronym for rearrangement induced premeiotically was first discovered in *N. crassa* by Eric Selker and Judith Stevens in 1986 and renamed repeat-induced point mutation (Cambareri et al. 1989) (Figure 1.3b). RIP detects duplicated sequences and introduces G:C to A:T transition mutations. RIP-affected sequences are prone to DNA methylation during the dikaryotic stage, methylated sequences are not limited to CpG dinucleotides, and cytosines are mutated on both strands of each DNA duplex (Singer et al. 1995). Two possible mechanisms have been hypothesized. Either a RIP mutation could result from enzymatic deamination of cytosine to form uracil followed by DNA replication or DNA repair or, alternatively, the mutation could result from cytosine methylation catalyzed by a specific cytosine DNA methyltransferase followed by deamination (Figure 1.3) (Selker 1990). However, so far, no experimental evidence has confirmed either of the above hypotheses. To date the only characterized gene involved in RIP is *rid* (RIP- defective).

In *N. crassa*, mutations in the *rid* gene, encoding a putative DNA methyltransferase, result in fertile RIP defective mutants (Freitag et al., 2002). RIP has also been observed in other ascomycete species. Experimental evidence of functional RIP is seen in *Magnaporthe oryzae* (Dean et al. 2005; Ikeda et al. 2002), *Podospora anserina* (Graia et al. 2001), *Leptosphaeria maculans* (Idnurm and Howlett 2003), *Fusarium graminearum* (Cuomo et al. 2007), and *Nectria haematococca* (Coleman et al. 2009). *In silico* signatures, viz. genomic sequences showing C:G to T:A transitions mostly via sequencing of TEs, have been reported for *M. oryzae*, *A. fumigatus*, *A. nidulans*, *A. niger*, *Fusarium oxysporum* and *Penicillium chrysogenum* (Malagnac and Silar 2011). RID homologs such as MASC1 in *Ascobolus immersus* (Malagnac et al. 1997), DmtA in *A. nidulans* (Freitag et al. 2002) and RID in *Sordaria macrospora* carry out functions independent of RIP (Nowrousian et al. 2010).

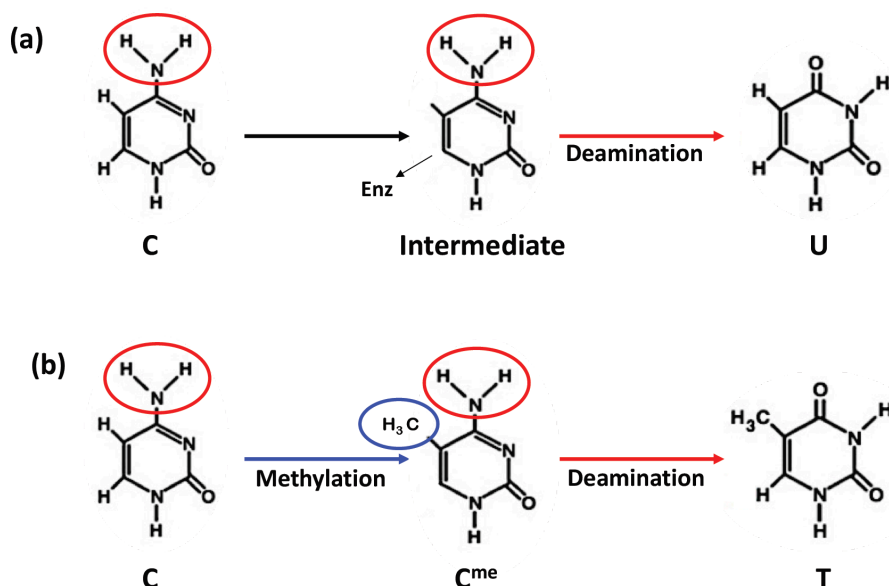


Figure 1.3 Possible mechanisms for C to T mutation by RIP.

1.4.1.2 MIP

MIP (Methylation Induced Premeiotically), first described in *Ascobolus immersus* is similar to RIP, but silencing is mediated by cytosine methylation alone and does not involve transition mutations in the DNA sequence. In *A. immersus*, the *mascl* (Methyltransferase from *Ascobolus* 1) gene, encoding a putative DNA methyltransferase, is responsible for the *de novo* MIP process (Malagnac et al. 1997). Like RIP, MIP also

appears to detect repeats involving either native or foreign sequences in a pairwise fashion (Faugeron et al. 1990; Fincham et al. 1989). *De novo* methylation of repeated sequences occurs prior to the sexual cycle (Rossignol and Picard 1991; Rhounim et al. 1992). Unlike RIP mutations, the silenced phenotype could be reversed during subsequent vegetative growth by the loss of cytosine methylation (Gladyshev 2017).

1.4.2 Post-transcriptional gene silencing

Proteins involved in PTGS mechanism are present in all lineages of eukaryotes (Cavalier-Smith, 2009; Cerutti and Casas-Mollano, 2006). In *N. crassa* two PTGS mechanisms have been studied so far: quelling and meiotic silencing of unpaired DNA (MSUD).

1.4.2.1 Quelling

Quelling, the best-characterized gene silencing mechanism in *N. crassa*, was first described in 1992 by Romano and Macino when they were attempting to overexpress the albino-1 (*al-1*) gene that is involved in carotenoid biosynthesis and orange pigmentation in *N. crassa*. The introduction of extra copies of the *al-1* gene gave rise to colonies of a white phenotype, reminiscent of the *al-1* mutant phenotype (Fulci and Macino, 2007; Pickford et al., 2002). It was found that the overexpressed transgene induces gene silencing at the posttranscriptional level, and the phenomenon was termed 'quelling' (Fulci and Macino, 2007). This process depends on the typical RNA-interference machinery. The model for quelling in *N. crassa* (Figure 1.4a) is similar to the models for RNAi in other organisms (Pickford et al., 2002). The proteins involved include QDE-1, a presumed RNA-dependent RNA polymerase (RdRP) (Cogoni and Macino, 1999a) which recognizes an aberrant RNA (aRNA) and converts it into a double strand RNA (dsRNA). The dsRNA are degraded by DCL (dicer-like) proteins into 20–25 nucleotide interfering RNAs (Catalanotto et al., 2004). An Argonaute protein, QDE-2 that encodes a Piwi-PAZ domain, along with small interfering RNAs (siRNAs) becomes integrated into the RNA-induced silencing complex (RISC) (Cerutti et al. 2000; Catalanotto et al. 2000). One of the siRNA strands is retained, QDE-2 then nicks the other siRNA strand, and it is removed by QIP (QDE-2 interacting protein) exonuclease (Maiti et al., 2007). The siRNA, which remains associated with RISC is guided to homologous mRNAs, which when trapped are most likely degraded by QDE-2. QDE-3, a RecQ DNA helicase, is thought to be involved in nucleic acid pairing (Catalanotto et al., 2004).

1.4.2.2 MSUD

Meiotic silencing by unpaired DNA is a PTGS mechanism also present in *N. crassa*. It was first described as 'meiotic transvection' and is specifically active during meiosis after karyogamy (Aramayo and Metzenberg, 1996). MSUD acts as a defense mechanism and protects the host genome against transposons and viruses by transiently silencing expression of genes homologues to the unpaired DNA that remains during meiotic pairing and recombination (Shiu et al., 2001). The general mechanism of quelling and MSUD are similar except for the differences in the sets of proteins that are involved (Galagan et al., 2003) (Figure 1.4c). The model consists of an unpaired DNA that triggers the production of aberrant RNAs (aRNAs; abnormal RNA that enter the silencing pathway (Decker et al. 2015)) and induces mRNA degradation (Wassenegger and Pelissier 1998). The process involves SAD-1/SAD-3, that encodes a RdRP similar to QDE-1, which converts the aRNA to dsRNA and is processed into siRNA by DCL-1 (SMS-3). Subsequently, QIP/SMS-2 (suppressor of meiotic silencing) destroys the complementary mRNA transcribed from the aRNA template (Hammond et al. 2011).

The present study describes the identification of genes in *P. membranacea* encoding homologues to proteins involved in gene silencing mechanisms such as quelling and MSUD (II).

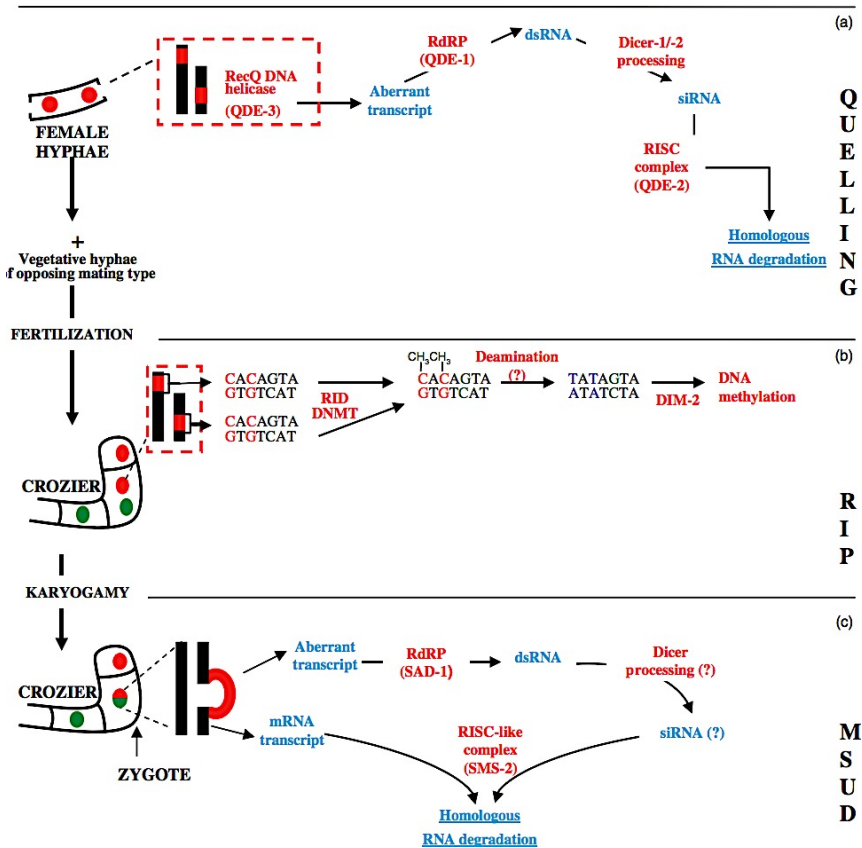


Figure 1.4 Gene silencing pathways in *Neurospora*.

(a) Quelling occurs in the vegetative phase when *Neurospora* grows as a haploid mycelium. In conditions of nitrogen starvation, each mating type (A and a) initiates the sexual phase and is fertilized by the opposite mating type. Fertilization begins with the cytoplasmic fusion between a specialized female hypha and a male vegetative cell. (b) During plasmogamy, the haploid nuclei are confined in a specialized hook-shaped structure (crozier), whereas repeat-induced point mutation takes place on duplicated sequences. Red C:G pairs represent targets of methylation by RID (a putative DNA methyltransferase). Putative deamination of these pairs can lead to a T: A conversion (blue). Unmutated C:G pairs (shown in green) are subsequently targeted by DIM-2 DNA methylation. (c) After DNA replication, the fusion of the two nuclei (karyogamy) occurs with the formation of the zygote. In this cell, unpaired DNA sequences are identified and silenced by the meiotic silencing by unpaired DNA pathway. (Reproduced from Catalanotto, 2006).

1.5 Technological background

1.5.1 Transcriptome–RNA sequencing (RNA–Seq)

Next-generation sequencing (NGS), also known as high-throughput sequencing (HTS) (Holt and Jones 2008) is a technology that offers a fast and efficient way to generate genomic, transcriptomic and methylomic datasets (Alagna et al. 2009; Dassanayake et al. 2009; Parchman et al. 2010). RNA–Seq (RNA Sequencing), is cost-effective and has revolutionized functional research. The technology offers new approaches for global measurement of gene expression, and indirectly of protein content. In this method, cDNA that is reverse transcribed from mRNA is sequenced, yielding information on the transcripts, their abundance, and complexity (Mortazavi et al. 2008; Torres et al. 2008). In brief, using oligo dT, mRNA with polyA is enriched and converted into pieces of cDNA using random hexamer primers, reverse transcriptase, RNase H and DNA polymerase I. Adapters are then ligated to the ends of the cDNAs, which are then amplified, and nucleotide reads are generated (Figure 1.5). Generally, eukaryotic mRNAs are polyadenylated. In addition to the mycobiont mRNA, the cyanolichen *P. membranacea* transcriptome contains prokaryotic mRNA from the cyanobiont (*Nostoc* sp.) and lichen-associated bacteria, devoid of poly-A tails (Régnier 2000-2013).

Depending on the sequencing technologies, the length of the reads ranges from 35–400 bp. For RNA–Seq, second-generation sequencing technologies such as the Illumina/Solexa (Lister et al. 2008; Mortazavi et al. 2008; Nagalakshmi et al. 2008; Wilhelm et al. 2008), Ion Torrent (Mangul et al. 2014), Applied Biosystems SOLiD (Cloonan et al. 2008), Roche/454 pyrosequencing (Barbazuk et al. 2007; Emrich et al. 2007; Vera et al. 2008), and third-generation sequencing (also known as long-read sequencing) technologies such as the Pacific Biosciences (PacBio), Oxford Nanopore Technologies (ONT) systems can be used (Munroe and Harris 2010; Lee et al. 2016). The RNA–Seq reads generated are either aligned to a reference genome using tools such as BWA (Burrows-Wheel Aligner) (Li and Durbin 2010), BLAT (BLAST-Like Alignment Tool), or Bowtie (Langmead and Salzberg 2012; Langmead et al. 2009) followed by a spliced-read aligner like TopHat (Trapnell et al. 2009); or *de novo* assembled using different tools such as Trinity (Zhao et al. 2011; Grabherr et al. 2011), RSEM (Li and Dewey 2011), or EBARDenovo (Chu et al. 2013). After the read alignment, tools such as Seqmonk (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>),

Cufflinks (Trapnell et al. 2010), DESeq (Love et al. 2014) or edgeR (Robinson et al. 2010) are used to provide information on the number of genes expressed, the mRNA copy numbers and transcriptional features such as start, end, and splice sites of the transcripts and isoforms. Thus, differences between tissues and conditions can be studied using RNA-Seq by measuring the expression level of transcripts derived from different tissues types and conditions.

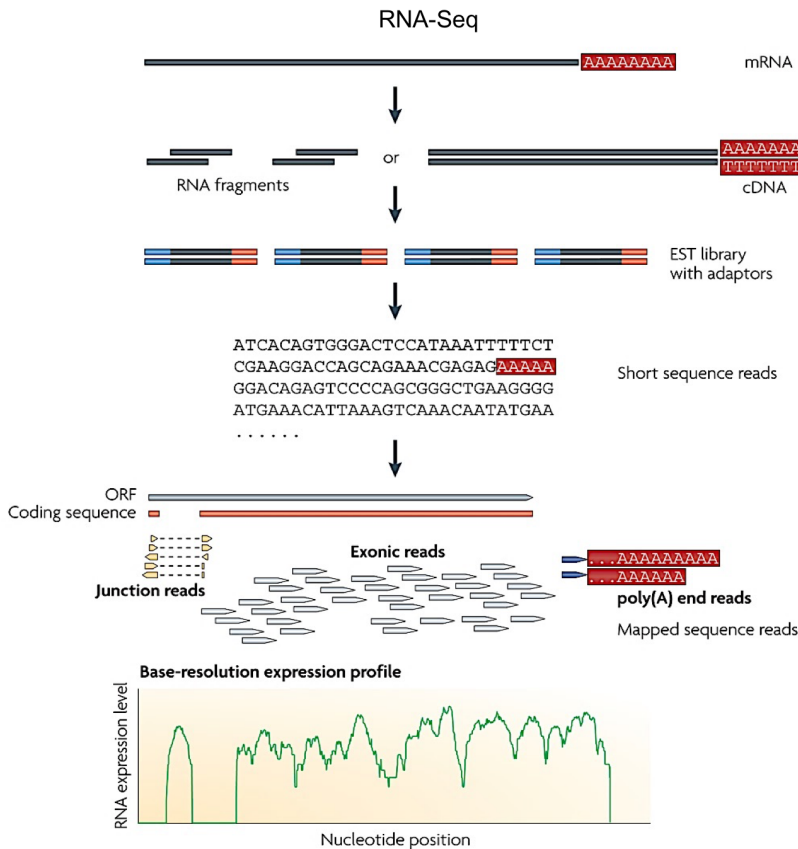


Figure 1.5 RNA-Seq.

Briefly, mRNAs are first converted into a library of cDNA fragments through fragmentation. Sequencing adaptors (blue) are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence reads are mapped to the reference genome or transcriptome. (Reproduced from Wang, 2009).

1.5.2 Methylome– Bisulfite sequencing (BS–Seq)

The cytosine methylome, the 5-methylcytosine (5mC) status of a genome at single-base resolution can be studied using bisulfite converted genomic DNA and NGS, i.e., BS–Seq. Frommer et al. in 1992 first introduced this method wherein bisulfite sulfonates unmethylated cytosines but not 5mC. The sulfonated cytosines are then deaminated and converted to uracil upon desulfonation (Figure 1.6). In brief, methylated adapters are ligated to sheared genomic DNA, it is denatured, and cytosines are converted into uracil using bisulfite, whereas 5mC is not susceptible to bisulfite and remains unconverted. Subsequent PCR amplification recognizes uracil as thymine and 5mCs that remain unconverted are recognized as cytosines and thus distinguished from the converted unmethylated cytosines. This is followed by either sequencing the PCR products directly (where average methylation status is detected) or cloning the PCR products and sequencing (where methylation patterns in single molecules are detected) (Figure 1.7). The BS–Seq reads that are generated are aligned to a reference genome using different tools such as BRAT (Harris et al. 2010), BSMAP (Xi and Li 2009), Bismark (Krueger and Andrews 2011), or BS–Seeker (Chen et al. 2010). The data provides information on the genome-wide methylation profiles of an organism at single-base resolution, and the roles of DNA methylation in biological processes can be further elucidated.

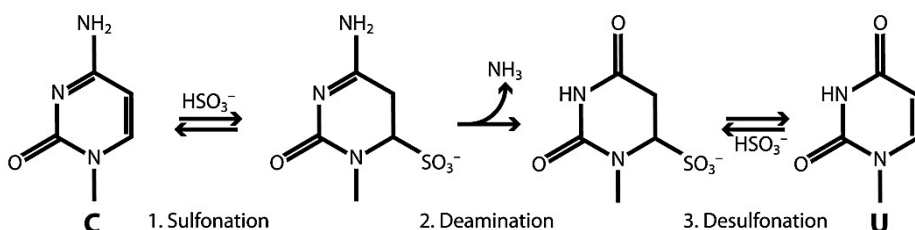


Figure 1.6 Schematic of Bisulfite conversion of Cytosine to Uracil.
(Reproduced from Patterson, 2011)

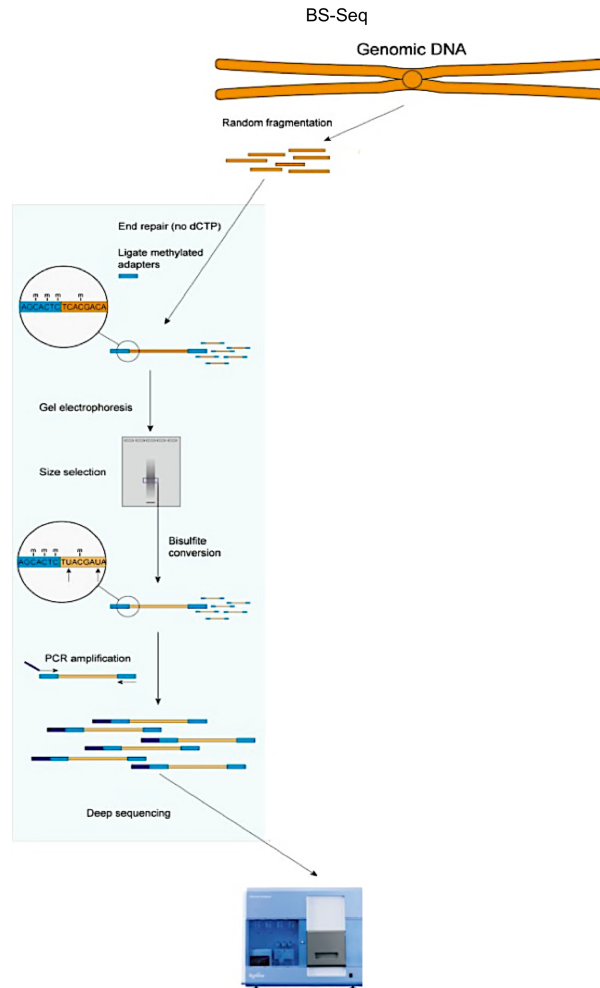


Figure 1.7 BS-Seq.

DNA is fragmented and double-stranded universal adapter sequences in which all cytosines are methylated are ligated. Treatment with sodium bisulfite converts unmethylated cytosines to thymine, followed by PCR enrichment of the libraries with primers complementary to the universal adapter sequences that can be sequenced. (Reproduced from Lister, 2009).

Thus, technological advances in sequencing have now made it possible to obtain global information on the DNA and RNA complement of organisms providing an interconnected database of the genome and transcriptome. Additionally, global epigenetic modifications such as DNA methylation patterns can also be characterized and compared to global gene expression profiles. To achieve this goal, RNA-Seq and BS-Seq

were used to study the transcriptome (I) and methylome (II), respectively of *P. membranacea*.

With the advent of 'omics' studies (see below) it is now possible to decode symbiosis at the molecular level. Studies involving assessment of tissue-specific gene expression (I) can enable us to make an inventory of genes that are specifically expressed in different lichen states, such as aposymbiotic and symbiotic tissues. The present study describes the tissue-specific transcriptomic landscape (I) and DNA cytosine methylation and genome defense (II) in the cyanolichen *P. membranacea*.

1.6 Omics in lichens

The methods of 'omics', which include genomics, proteomics, methylomics, and metabolomics, can be used to shed light on the workings of symbiotic systems such as lichens. NGS has been implemented to study aspects of the lichen symbiosis wherein the identification of genes and metabolic pathways are transforming our understanding. Recently, Spribille et al. (2016) augmented the one lichen-one fungus concept by describing a second fungal partner in the lichen symbiosis. Based on the transcriptomics approach, basidiomycete yeasts were shown to be active in the cortex of some ascomycete macrolichens. Several other studies utilizing the omics approach on lichens are listed in Table 1.1.

Table 1.1 Omics studies on lichens.

Study organism	Study	References
<i>Peltigera membranacea</i> & <i>Peltigera malacea</i>	Characterization of the mitochondrial genomes	Xavier et al. (2012)
<i>Cladonia grayi</i>	Identification of horizontal transfer of ammonium transporter genes from prokaryotes to <i>C. grayi</i> and characterization of the DNA methylation status of the lichen	McDonald et al. (2012)
<i>Cladonia rangiferina</i>	Characterization of the transcriptome and Discovery of genes involved in metabolic pathways	Junttila and Rudd (2012)
<i>Peltigera membranacea</i>	Identification of pedrin type biosynthetic gene cluster nsp involved in secondary metabolite synthesis	Kampa et al. (2013)
<i>Endocarpon pusillum</i>	Characterization of the genome	Wang et al. (2014a)
<i>Endocarpon pusillum</i>	Identification of genes involved in drought tolerance by comparative transcriptome analysis	Wang et al. (2015)
<i>Peltigera membranacea</i>	Characterization of bacteriomes of <i>P. membranacea</i> . Diverse non-phototrophic bacteria that are involved in phosphate solubilization and biopolymer degradation were identified	Sigurbjornsdottir et al. (2015)
<i>Lobaria pulmonaria</i>	Population genetics of lichen fungi	Werth et al. (2015)
<i>Lobaria pulmonaria</i>	Maintenance of lichen symbiosis by active contribution of microbiome	Cernava et al. (2017)
<i>Peltigera ponojensis</i>	Microscopic and metagenomic analyses of <i>Peltigera ponojensis</i> (Peltigerales, Ascomycota)	Graham et al. (2018)
<i>Cladonia grayi</i>	Publicly available whole genomes and ESTs or transcriptomes	https://genome.jgi.doe.gov/Clagr2/Clagr2.home.html
<i>Cladonia macilenta</i>		https://gold.jgi.doe.gov/projects?id=42523
<i>Lobaria pulmonaria</i>		https://genome.jgi.doe.gov/portal/LobpulMetagenome_10_FD/LobpulMetagenome_10_FD.info.html
<i>Trypethelium eluteriae</i>		http://genome.jgi.doe.gov/Tryvi1/Tryvi1.home.html
<i>Xanthoria parietina</i>		https://genome.jgi.doe.gov/Xanpa1/Xanpa1.home.html

1.6.1 Metagenomic snapshot of the *Peltigera membranacea* associated microbiome

Lichen thalli host diverse populations of organisms, such as lichenicolous (growing on the thallus) and endolichenic (growing within the thallus) fungal species (Bates et al. 2011; Arnold et al. 2009; Lawrey and Diederich 2003). Lichenicolous fungi often cause visible changes (e.g. galls) or produce characteristic fructifications and have been well known for a long time, but recent studies have revealed a Cystobasidiomycete yeast phase (*Cyphobasidium* yeasts, Pucciniomycotina) (Millanes et al. 2016) which may reside asymptotically in *Bryoria capillaris* and other lichens. Among the Peltigeralean lichens that were screened by Spribille et al., other Cystobasidiomycetes but not *Cyphobasidium* yeasts were identified either in *Peltigera*, *Lobaria* s. lat. or in the related genera (Spribille et al. 2016). *Nostoc* spp. are considered to be the primary photobiont in most cyanolichens, but recent studies have shown that in addition to secondary fungi, lichens host diverse assemblages from several bacterial phyla (Bates et al. 2011; Hodkinson and Lutzoni 2009; Grube et al. 2009; Arnold et al. 2009; Cardinale et al. 2008; Liba et al. 2006; Cardinale et al. 2006; Gonzalez et al. 2005; Sigurbjornsdottir et al. 2015).

A reference-based binning approach was carried out to search for yeasts or other fungi in the *P. membranacea* metagenome data. Roche 454 pyrosequencing reads were mapped using CLC Genomics Workbench (v9.5.3) to (1) *P. membranacea* internal transcribed spacer (ITS) regions (2) *Cyphobasidium usneicola* ITS regions. Furthermore, operational taxonomic units (OTUs) analysis was undertaken by clustering the reads into OTUs using a 98% similarity threshold. The *P. membranacea*-prokaryotic consortium was also examined, and the results are presented below. *P. membranacea* sample collection, genomic DNA extraction, and metagenomic analysis were performed as described in Appendix A and B. The number of reads generated for each tissue is provided in Table S2.

The study presented here provides a snapshot of the *Peltigera membranacea* microbiome. Firstly, the reads from apothecia and thallus were mapped to *P. membranacea* ITS1-5.8S-ITS2 regions (759 bp, accession no: JX181776) using relaxed parameters (Figure 1.8). To facilitate mapping of reads from other mycobionts if any, less stringent mapping at a similarity threshold of 80% that enables mapping of higher

number of reads and allows indels were used for further analysis. Mapping of reads at similarities 80%, 90%, 98% and 99% are listed in Table 1.2). BLASTN analysis against nr database for reads >200 bases in length (195 reads from thallus and 445 reads from apothecia) were performed, and all the hits were to seven accession numbers of *P. membranacea* ITS sequences (Table 1.3). Secondly, the reads were mapped to *C. usneicola* ITS1-5.8S-ITS2 regions (501 bp, accession no: KU587702) using the above parameters (Figure 1.9). At 80% similarity, 76 and 31 reads from apothecia and thallus, respectively mapped to the 501 bp region. All the reads mapped to the highly conserved 5.8S rRNA region (Figure 1.9). BLASTN analysis of the reads that were >70 bases in length (total 90) were carried out and all the reads except two showed hits to *P. membranacea* ITS sequences. (Table 1.3). Out of 90 reads, two reads less than 100 bp were to *P. canina*.



Figure 1.8 Mapping of reads to *P. membranacea* ITS1-5.8S-ITS2 region. Mapping of reads from A. apothecia B. thallus. ITS1, 5.8S and ITS2 regions are depicted. Forward and reverse reads are shown as green and red colors, respectively.

Table 1.2 Summary of reads mapped.

454 reads	Similarity fraction	Total reads	Mapped reads	Not mapped reads
Thallus whole	0.80	2,450,573	268	2,450,305
	0.90		219	2,450,354
	0.98		170	2,450,403
	0.99		133	2,450,440
Apothecia	0.80	2,240,326	613	2,239,713
	0.90		494	2,239,832
	0.98		359	2,239,967
	0.99		256	2,240,070

Table 1.3: Summary of BLASTN hits of 454 reads.

ITS sequences	Accessions	Organism	Apothecial reads	Thallus reads
<i>Peltigera membranacea</i>	FJ709035	<i>Peltigera membranacea</i>	2	3
	JX181776		313	118
	JX195294		1	1
	JX195297		7	5
	LT852896		1	1
	LT852898		73	26
	LT852899		48	41
<i>Cyphobasidium usneicola</i>	AY257908	<i>Peltigera membranacea</i>	1	1
	LT852898		7	3
	LT852899		52	22
	JX181776		1	0
	LT852826	<i>Peltigera canina</i>	2	1

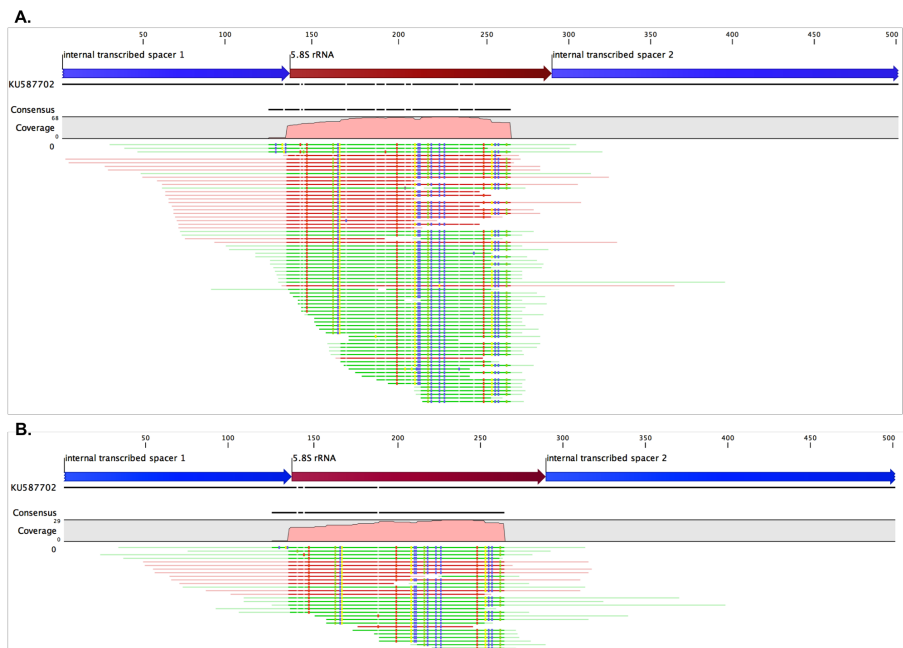


Figure 1.9 Mapping of reads to *Cyphobasidium usneicola* ITS1-5.8S-ITS2 regions. Mapping of reads from A. apothecia B. thallus. ITS1, 5.8S and ITS2 regions are depicted. Forward and reverse reads are shown as green and red colors, respectively.

Lastly, taxonomic assignments of OTUs using the UNITE (User friendly Nordic ITS Ectomycorrhiza) database (<https://unite.ut.ee/>) (Kõljalg et al. 2014) for ITS (fungi), and the SILVA database for 16S, 18S and 28S rRNA were carried out (Appendix B1). The fungal ITS based OTU analysis at 97% similarity showed an exclusive dominance of *Peltigera membranacea* (100%) for both apothecia and thallus samples. For 16S rDNA subunit, taxonomic assignments of bacterial OTUs at the were dominated by Proteobacteria and *Nostoc* sp. PCC73102 (Table 1.4, Figure S1 and S2). The results were in agreement with the findings of Sigurbjornsdottir et al. (2015) where the bacteriome was determined using the *P. membranacea* 454allcontigs database. Additional studies are needed to determine if there is variation in the *Peltigera*-associated microbiome directly linked with geographical location and/or chemical/physical gradients of the environment.

For 18S rDNA subunit, a total of 46 and 12 reads were classified as *Peltigera* and *Solorina* OTUs, respectively. Similarly, for 28S rDNA subunit, 239 and 12 reads were assigned to Peltigerales and Umbilicariales, respectively (Table 1.2). The same result was observed in a similar study using *Peltigera ponojensis* where 11% of 18S rDNA reads were classified as *Solorina*, and no Pucciniomycotina was detected (Graham et al. 2018). To confirm the presence or absence of *Solorina* mycobiont in the *P. membranacea* metagenome, the 454 reads from apothecia and thallus were further analysed (see Appendix B2 for details). Upon further investigation, it is confirmed that there is no indication of additional mycobiont such as *Solorina* in *P. membranacea* metagenome. Henceforth, results should be interpreted with caution and should be validated by additional analysis as undertaken in this study. From the above analysis it is evident that *Cyphobasidium* spp. or other Cystobasidiomycetes are not present in *P. membranacea* metagenome at similar levels found in a number of lichens by Spribille et al. (2016). Also, no Pucciniomycotina was detected (Table 1.4.).

Table 1.4: Taxonomic assignments of 454 reads using the SILVA database (v132).

Apothecia 454 reads	
Total	11407
Classified	352
No Relative	11055
Bacteria	86
<i>Nostoc</i> PCC 73102	6
Proteobacteria	47
Bacteroidia	17
Actinobacteria	10
Bacillus	1
Blastocatellaceae	2
Chthoniobacterales	3
Eukaryota	266
Unassigned eukaryota	1
Archaeplastida	4
Opisthokonta	261
Bilateria	3
Ascomycota	258
Eurotiales	3
Pezizomycotina	255
Lecanoromycetes	251
Peltigerales	239
Umbilicariales	12
Thallus 454 reads	
Total	247
Classified	173
No Relative	74
Bacteria	106
<i>Nostoc</i> PCC 73102	45
Proteobacteria	53
Alphaproteobacteria	50
Gammaproteobacteria	3
Actinobacteria	3
Eukaryota	67
Alveolata	2
Opisthokonta	65
Pezizomycotina	64
Lecanoromycetes	60
Peltigeracea	58
Peltigera	46
Solorina	12
Bdelloidea	1

1.7 Objectives

The central rationale of this thesis is the macromolecular characterization of the mycobiont present in the cyanolichen *Peltigera membranacea* with the aim of gaining basic knowledge about the lichen symbionts and their interrelationship, and how they interact with the natural environment. The principal tasks were to study the transcriptomics (I) and methylomics (II) of the cyanolichen *P. membranacea* to provide a strong informational basis for a functional understanding of the symbiosis. It was first necessary to annotate (I) the *de novo* assembled mycobiont genome, which was available for this lichen. To this end, *de novo* transcriptomes (I) derived from the symbiotic (thallus) and aposymbiotic (apothecia and rhizine) tissues of *P. membranacea* were used. For the transcriptomics study (I), quantitative Illumina RNA-Seq data from each of the tissues (apothecia, rhizines and thallus) were used to identify tissue-specific gene expression potentially involved in fundamental development/specialization and metabolic processes of the lichen as a whole.

In fungi, genome defense mechanisms include DNA methylation, repeat induced point mutation (RIP), methylation induced premeiotically (MIP) and RNA silencing (quelling and meiotic silencing of unpaired DNA (MSUD)). Within this context, the aim was to find possible candidate genes involved in the above genome defense mechanisms. Also, identification of genes involved in nutrient assimilation, symbiosis, DNA methylation, RNA-silencing, reproduction, melanin and glyoxylate pathways was carried out (I). Work was focused on galectin-like genes and their proteins which are thought to play a role in recognition of potential partners in the establishment of the lichen. *lec-1* (III) and *lec-2* (IV) were identified and characterized. Furthermore, for the methylomics study (II), the objective was to obtain BS-Seq data for the thallus and rhizine tissues using Illumina sequencing and determine the genome-wide, single nucleotide resolution maps of methylated cytosines in the mycobiont DNA to better understand genome regulation and genome defenses. Auxiliary to the primary objectives, a lichen which was collected during field sampling was further investigated and identified as a new lichen species, *Peltigera islandica* (V).

1.8 Overview of chapters

Chapter 1, "Introduction" starts with a background to lichens and the model lichen *Peltigera membranacea* used in the study. The anatomy and morphology of the three tissues of *P. membranacea*, apothecia (aposymbiotic fungal ascocarp), thallus (symbiotic), and rhizines (aposymbiotic hairlike vegetative fungal structures) are described. The symbiotic genes of lichens along with the different gene silencing mechanisms in fungi are briefly presented. Nucleic acid sequencing methods such as RNA-Seq and BS-Seq used to study the transcriptome and methylome, respectively, of *P. membranacea* tissues, are described. A table showing the application of omics technologies in lichens together with a metagenomic snapshot of *P. membranacea* is presented in this chapter.

Chapter 2, "The tissue-specific transcriptomic landscape of the lichen *Peltigera membranacea*" discusses the isolation of RNA and the sequencing of mycobiont poly(A) mRNA from three lichen tissues; apothecia, rhizines, and thallus. Here, *P. membranacea* poly(A) mRNA was subjected to in-depth transcriptome analysis to identify differentially expressed genes. Using Illumina sequencing, roughly 600 million reads were obtained from the three tissues. The RNA-Seq reads were *de novo* assembled into transcripts and were used as expressed sequence tags (ESTs) to annotate the genome using the AUGUSTUS software. Genes were identified and assigned functionality. 16,400 protein-coding genes were predicted, and one or multiple GO terms were assigned. 112 specific KEGG pathways were predicted. Genes putatively involved in nutrient assimilation, as well as genes of the glyoxylate and melanin pathways were identified and are reported along with their expression levels. Genes thought to be associated with symbiosis, sexual reproduction and mating-type genes were also identified.

Chapter 3, "DNA cytosine methylation and genome defense in the lichen *Peltigera membranacea*" addresses DNA methylation and genome defense in this lichen. DNA methylation is a fundamental biological process essential for the normal development of mammals and plants and is associated with key processes such as silencing of transposons, regulation of genes and cell differentiation. For single-base resolution mapping of DNA methylation, Illumina sequencing of the bisulfite converted DNA was performed on metagenomic DNA isolated from rhizine and thallus tissues. Cytosine methylation was identified in all

three sequence contexts (CG, CHG and CHH) and methylation was primarily found in repetitive elements. Methylation in transposons (TE) showed an inverse correlation to expression whereas differential exon methylation between the tissues (thallus and rhizines) showed no correlation to expression. The *rid* gene in *N. crassa*, which is known to be required for RIP (Repeat-induced point mutation) was absent from the *P. membranacea* genome, indicating that it lacks the ability for RIPs. Homologues of *N. crassa* methylation and RNA silencing genes were identified in the *P. membranacea* genome.

Chapter 4, "Galectin-like genes: Identification and expression of *lec-1* and *lec-2* in *Peltigera membranacea*" describes two genes encoding galectin-like proteins in the lichen mycobiont. Analysis of the gene structures indicated that both were derived from the mycobiont and this was confirmed by transcriptome analysis of freshly collected samples in which expression of *lec-1* and *lec-2* were detected in apothecia and rhizines, tissues that are composed of only fungal cells. A comparative study showed that expression of *lec-1* in portions of the tissues that included both symbionts was up to two orders of magnitude higher than in tissues lacking photobionts, consistent with a model in which the recognition of photobiont ligands by lectins such as *lec-1* is a factor in the establishment or maintenance of compatible symbionts. Expression of *lec-2* was not correlated with the presence of photobionts and was reduced only in rhizines. Sequencing of *lec-2* open reading frames from 21 individual samples showed an unexpectedly high level of polymorphism in the deduced protein (LEC-2), which was sorted into nine haplotypes, based on amino acid sequence. The rates of nonsynonymous versus synonymous nucleotide substitutions deviated significantly from the null hypothesis of neutrality, indicating strong positive selection. Molecular modeling revealed that most amino acid replacements were around the putative carbohydrate-binding pocket, indicating changes in ligand binding. Lectins have been thought to be involved in recognition of photobiont partners in lichen symbioses, and the hypothesis that positive selection of LEC-2 is driven by variation in the *Nostoc* photobiont partner was tested by comparing mycobiont LEC-2 haplotypes and photobiont genotypes, as represented by the *rbcLX* region. It was not possible to pair up the two types of marker sequences without lineage conflicts, suggesting that positive selection of LEC-2 was not due to variation in photobiont partners.

Chapter 5, "*Peltigera islandica*, a new cyanolichen species in section *Peltigera* ('*P. canina* group') ", describes the identification of a new species. This species is similar in general appearance to *P. rufescens* and *P. membranacea* but may be recognized by its downturned lobe tips and narrow lobes. Most thalli are bright emerald green in colour when moist, although a dark khaki green colour-morph is also documented. Monophyly of *P. islandica* s. lat. (i.e., including *Peltigera* sp. A, sensu O'Brien et al., from Canada) is significantly supported based on ITS sequences and corroborated by molecular synapomorphy (absence of the ITS1 hypervariable region). Analysis of the *rbcLX* locus indicates that cyanobionts of *P. islandica* (*Nostoc* sp.) are derived from a pool of Icelandic genotypes, some of which are present in other *Peltigera* species, including *P. 'neorufescens'*, another taxon new to Iceland collected during this study. Association with photobionts that are shared by other local species suggests *P. islandica* may be well established in Iceland, but a review of herbarium collections, as well as broader field surveys, are needed to characterize its geographical distribution better.

CHAPTER 2

THE TISSUE–SPECIFIC TRANSCRIPTOMIC LANDSCAPE OF THE LICHEN *PELTIGERA* *MEMBRANACEA*

2 THE TISSUE–SPECIFIC TRANSCRIPTOMIC LANDSCAPE OF THE LICHEN *PELTIGERA MEMBRANACEA*

2.1 Introduction

Peltigera (Peltigerineae, lichenized Ascomycota) is one of the most widespread lichen genera and contains bipartite and tripartite species (Miadlikowska and Lutzoni 2004). Bipartite *Peltigera* have *Nostoc* spp. Cyanobacteria as their primary photobiont and tripartite *Peltigera* include both *Coccomyxa* sp. green alga and *Nostoc* spp., the latter forming external cephalodia. These lichens commonly occur in humid or moist, shaded habitats, often on the forest floor, or along roadsides and in other disturbed habitats (Richardson 1992; Park et al. 2014; Langmead et al. 2010).

Lichens have been successfully studied using transcriptomic and genomic approaches. Over the past few years, efforts have been made to understand the lichen symbiosis and elucidate molecular mechanisms through next-generation sequencing (NGS) approaches. Genome–wide expression patterns have been studied for the chlorolichens *Cladonia grayi* (Joneson et al. 2011; Junttila and Rudd 2012; Pal et al. 2007; McDonald et al. 2012; Calmes et al. 2013; Collins and Farrar 1978), *Cladonia rangiferina* (Junttila and Rudd 2012), *Endocarpon pusillum* (Wang et al. 2014) and *Lobaria pulmonaria* (Werth and Andrésson, unpublished). The annotated genomes of *Xanthoria parietina* (<https://genome.jgi.doe.gov/Xanpa1/Xanpa1.home.html>), *Cladonia grayi* (<https://genome.jgi.doe.gov/Clagr3/Clagr3.info.html>) and *Lobaria pulmonaria* (<https://genome.jgi.doe.gov/Lobpul1/Lobpul1.home.html>) are accessible from the JGI portal (Grigoriev et al. 2012). Genomic and transcriptomic data are available for cultured mycobiont and photobiont as well as the thallus tissues, but only for chlorolichens.

Very few lichens have been successfully cultivated in the laboratory. Attempts to grow mycobionts or reconstitute lichens such as *P. membranacea* have proven to be difficult with limited success as it is difficult to create a similar niche in the laboratory as found in nature for symbiosis to take place between fungi and cyanobacteria. Therefore, direct extraction of RNA from field samples and sequencing are used for metatranscriptomic studies. Ideally, sequencing of RNA transcripts from field samples can provide an unbiased approach in investigating the complexity and the diverse physiologies of uncultured organisms in their natural habitat (Bomar et al. 2011). In theory, metatranscriptomics, where total RNA from the environment is sequenced, can reveal active community members and functional metabolic pathways in the natural ecosystem (Wisecaver et al. 2014).

RNA-Seq is a practical approach to study gene expression in all species including non-model species, where genetic tools, and genomic sequences and resources are limited. In general, tissue-specific expression studies are experimentally challenging in non-model organisms such as lichens, as separation of tissues for RNA isolation is a time-consuming and laborious task (e.g., separation of rhizines from thallus in *P. membranacea*), which can lead to RNA deterioration and artifacts. Moreover, isolation of RNA from lichens is challenging as they have rigid cell walls, are rich in polysaccharides, lipids, exoenzymes, and microbes that interfere with isolation or degrade RNA, restricting the quality and yield of the RNA. Also, organisms that are recalcitrant to cultivation in the laboratory present severe limitations to studies as their physiology, metabolic activities and various characters cannot be readily manipulated.

Tissue-specific gene expression plays a fundamental role in multicellular biology and provides key biological information on the tissue types within an organism (Jensen et al. 1991). Large-scale gene expression data obtained with current sequencing technologies can provide measurements of transcript abundance in various tissues, thus offering the ability to understand nuanced aspects of symbiosis in *P. membranacea* by identifying genes differentially expressed in the symbiotic (thallus) and aposymbiotic (apothecia and rhizine) tissues. The *P. membranacea* cyanolichen is large enough to allow the dissection of individual tissues, thereby enabling to study tissue-specific gene expression that will aid in elucidating the molecular mechanisms of tissue development, regulation and function of biological processes in this

lichen. In this study, the available genome sequence of *P. membranacea* (Xavier et al. 2012) is used to provide a genome-wide inventory of differential gene expression in three tissues (apothecia, rhizines and thallus) using RNA-Seq.

The lichen *Peltigera membranacea* has been sequenced at the University of Iceland using both 454 and Illumina technologies (Xavier et al. 2012) and has an estimated genome size of 43 Mbp. The genome sequence is an organism's blueprint, and annotation of the protein-coding genes in genomic sequences can provide useful insights into core functions of an organism (Salem et al. 2015). In eukaryotes, genome annotation which includes gene prediction is a complex endeavour (Li et al. 2011). The annotation is usually carried out by integrating multiple pieces of information (Haas et al. 2005), such as complementary DNAs (cDNAs) that represents the integral sequences of messenger RNAs (mRNAs) i.e. expressed sequence tags (ESTs), homology to proteins from closely related species and *de novo* predictions (Brent 2008). The basic assembly of the *P. membranacea* genome is not annotated, and therefore a transcriptome assisted genome annotation was carried out. The unannotated first draft assembly of *P. membranacea* is referred to as P.mem v1.0 and the annotated assembly as P.mem v1.1 (Appendix A3). Although far from perfect, the availability of annotated genes provides an important reference for this and future studies.

The four main objectives were to: (1) assess the completeness of the genome assembly and validate the assembly using transcriptomic data (2) generate a *de novo* transcriptome for gene identification in order to improve the accuracy of *de novo* gene prediction (3) assign functionality to the gene models, and (4) identify tissue-specific gene expression in the three different tissues (apothecia, thallus and rhizines), in order to elucidate the functions likely to be involved in the *P. membranacea* symbiosis. Genes of special interest, i.e. genes potentially involved in metabolism and assimilation of carbon and nitrogen, the melanin pathway, symbiosis, genome defense (see Chapter 3), and sexual and vegetative incompatibility related genes were studied to a greater depth.

2.2 Materials and Methods

2.2.1 Lichen Material and RNA extraction

Lichens were collected from Keldur, Reykjavik, Iceland. RNA was isolated from lichen tissue collected in RNAlater (Ambion), and the tissues were separated while immersed in RNAlater. Samples were treated within 24 hours of collection to minimize RNA degradation. The lichen tissue samples were finely ground in liquid nitrogen using a mortar and pestle. Several RNA extraction methods were tested, and these included (i) conventional phenol-chloroform extraction using LiCl precipitation (ii) TRI reagent (iii) RNAqueous®–Midi Total RNA Isolation Kit (Ambion™) and (vi) ToTALLY RNA™ Total RNA Isolation Kit (Ambion™). The ToTALLY RNA Kit was found suitable and was used to extract the RNA, according to manufacturer's instructions. Total RNA was isolated from 13 tissue samples (Table 2.2). The RNA isolated was run on a 1% agarose denaturing gel (Figure 2.1). Before library preparation, each sample was subjected to quality control as follows. The concentration and purity of RNA were assessed by a NanoDrop spectrophotometer (Thermo Scientific). Each sample was required to have an A260/A280 ratio between 2.0 and 2.2 and an A260/A230 ratio above 2.0. RNA integrity was evaluated using the RNA Nano Chip, Agilent 2100 Bioanalyzer (Agilent Technology Inc.), (Figure S5).

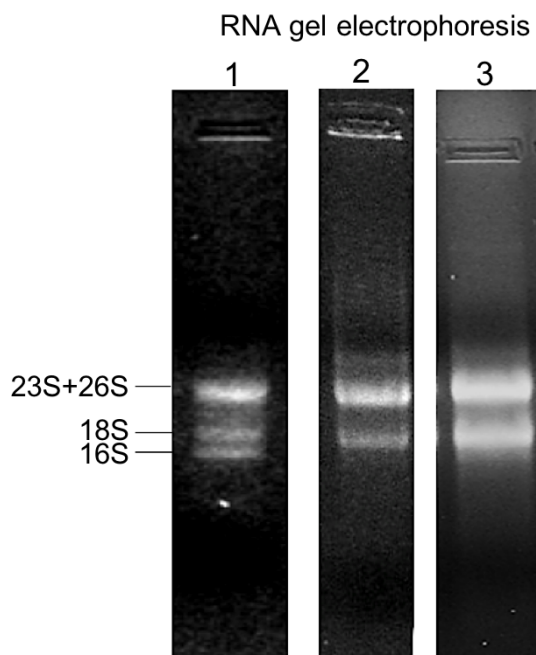


Figure 2.1 Denaturing RNA gel electrophoresis of total RNA. Lane 1: RNA isolated from thalli, lane 2: RNA from apothecia and lane 3: RNA from rhizines. 23S+26S, 18S and 16S ribosomal RNA (rRNA) bands are labeled. Note that the 16S band is absent from lanes 2 and 3.

2.2.2 RNA–Seq library preparation

mRNA was purified from 8–9 μg of total RNA from each tissue sample. The samples used in this study are listed in Table 2.2. A standard Illumina mRNA sequencing protocol was used for the tissues collected in 2009, 2010 and 2011. Multiplex sequencing was done on the tissues collected in 2012 using the TruSeqTM RNA protocol from Illumina. The standard workflow is as follows: RNAs are extracted, followed by purification of poly-A containing mRNA molecules using oligo-dT attached magnetic beads. The mRNA is subjected to fragmentation using divalent cations under elevated temperature. Following fragmentation, the cleaved RNA fragments are copied into first strand cDNA using reverse–transcriptase and random primers. Second strand cDNA is synthesised, using DNA Polymerase I and RNase H. The ends of the cDNA are repaired by the addition of a single ‘A’ base, followed by adapter ligation. The products are then purified and enriched with PCR to create the final cDNA library. After subsequent cluster generation, the cDNA libraries are subjected to

high-throughput sequencing. The mRNA libraries were sequenced on a Illumina Genome Analyzer at the deCODE Genetics facility (Reykjavik, Iceland), generating paired-end (PE) reads with 35, 85, 100 and 110 bases for tissues collected in 2009, 2010, 2012 and 2011, respectively.

2.2.3 RNA–Seq reads processing

The steps carried out to generate and analyse RNA–Seq data are shown in Figure 2.2. The raw RNA–Seq reads were trimmed of adapters and reads with a quality score < 0.05 were also trimmed using the 'Trim Sequences' tool implemented in the CLC Genomics Workbench v.9.5.1. Trimmed reads < 20 bp were discarded. It is debatable whether duplicate reads, i.e., reads that map to an identical genomic sequence should be removed or retained for RNA–Seq analysis. These reads stem from PCR duplication or identical RNA fragmentation during library preparation (Sendler et al. 2011). However, for the present study, the probable PCR duplicate reads were removed from the trimmed reads with the 'Duplicate Read Removal tool' from CLC Genomics Workbench v9.5.1. After trimming and duplicate removal, the processed RNA–Seq reads were checked using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to assess the quality of the reads and these processed reads were used in further analysis.

2.2.4 Generation of gene models, gene prediction and annotation, and metabolic pathway identification

Processed RNA–Seq reads from apothecia (autumn 2009), rhizines (autumn 2010) and thallus whole (autumn 2010 and summer 2011) tissues were *de novo* assembled using Trinity (Grabherr et al. 2011) (Figure 2.2). The P.mem v1.0 genome assembly was annotated using the evidence-based annotation tool, AUGUSTUS (Stanke et al. 2004) (Figure 2.2). First, AUGUSTUS was trained using *Aspergillus nidulans* as the reference (<http://fungus.genome.duke.edu/>). Second, transcripts generated from the three tissues using Trinity were combined into one unified, reference EST/mRNA base to create intron evidence, and for defining 5' and/or 3' UTRs in gene models. Third, the *de novo* transcriptomes and the annotated genes were compared by performing a BLASTN search between the two datasets. The isoforms of the annotated genes are suffixed as gene1.t1, gene1.t2, gene1.t3, etc. For further analysis, genes with prefix t1 from all the gene models were extracted, and a database was created. Each of the *de novo* transcriptomes were used as a query for

BLASTN searches, and the BLAST analysis was carried out. Fourth, protein sequences were derived from translations of CDS using the 'Translate CDS/ORF' option implemented in CLC Genomics Workbench. Finally, the annotated proteins were validated by aligning them against the curated proteins available from NCBI, and alignments were produced automatically using CLUSTALW available in CLC Genomics Workbench (Appendix F, Figure S7 – S20).

The Benchmarking Universal Single-Copy Orthologs (BUSCO) tool (Simao et al. 2015) was used for assessing the completeness of genomes and transcriptomes. The missing BUSCO protein sequences were retrieved from <http://www.orthodb.org/> using the EOG identifiers. Functional annotations were carried out for the AUGUSTUS derived proteins using the Blast2GO software (Gotz et al. 2008). The annotations included a best BLASTP match selection, Gene Ontology (GO) based data mining, Enzyme Code (EC) assignments and protein domain information from InterProScan (IPS). First, a BLASTP search against the NCBI nr database was done which was then followed by an InterProScan. GO mapping was done where all the functional information from GO database was assigned to the BLAST hits. After the prediction of GO terms, KO (Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology)) was assigned for each protein using BlastKOALA (KO and Links Annotation) and GhostKOALA (<http://www.kegg.jp/ghostkoala/>) (Kanehisa et al. 2016). KO identifiers (K numbers) generated from both BlastKOALA (3,171K) and GhostKOALA (3,573 K) were combined manually. K numbers were used as an input to reconstruct KEGG metabolic pathways (http://www.genome.jp/kegg/tool/map_pathway1.html), KEGG BRITE functional hierarchies (http://www.genome.jp/kegg/tool/map_brite.html) and KEGG modules (http://www.genome.jp/kegg/tool/map_module.html). The TMHMM plugin from the CLC Genomics Workbench v9.5.1 was used to predict trans-membrane proteins, and the annotations were added to the existing gff annotation file. BLASTP matches were filtered using an arbitrary cut-off of 1×10^{-6} and the gene descriptions generated from BLASTP were incorporated into the gff annotation file using a Perl script, `blast2go2gff.pl` (Simon Andrews, personal communication, May 23rd, 2014, Appendix M). The annotated version of the *P. membranacea* genome assembly is referred to as P.mem v1.1.

The millions of short *reads* generated were aligned to the *P. membranacea* reference genome and the transcript abundances measured.

However, the measurement can be biased, and RNA 3' biases in the different samples were analyzed using an R script (Jóhannes Guðbrandsson, personal communication, April 10th, 2013). Using the 'extract annotations' option in CLC Genomics Workbench v9.5.1 the parts of the sequence in the scaffolds that were labeled as transcripts were extracted which resulted in 5'-3' oriented sequences. The BAM files for the 5'-3' oriented sequences with RNA reads mapped onto the P.mem v1.1 genome were used as the input file in an R script. The ratio of reads mapped to the 3' half of the transcript compared the total length of the transcript was calculated, and a heat map was generated (Figure S6).

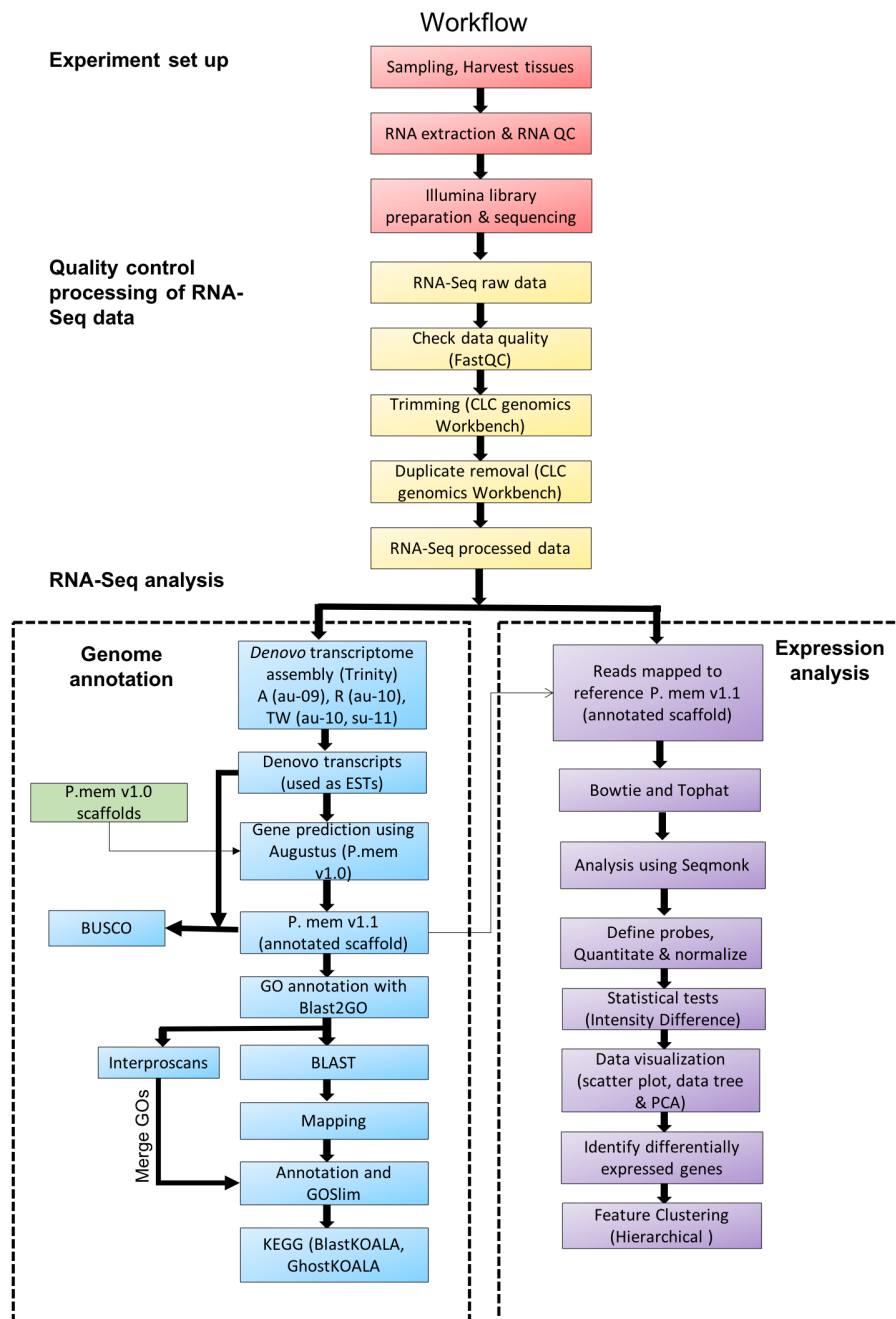


Figure 2.2 Genome annotation and RNA-Seq analysis workflow.

2.2.5 Identification and analysis of genes

The genes involved in symbiosis, metabolism and assimilation of carbon and nitrogen, the melanin pathway, the mating process, heterokaryon incompatibility, the DNA methylation and genome defense machinery were identified and annotated from P.mem v1.1 using the homologous genes from *Aspergillus nidulans*, *Aspergillus niger*, *Neurospora crassa* and *Podospora anserina* as queries. Additionally, two lectin-encoding genes, *lec-1* and *lec-2*, were described by comparison to published sequences from partial proteins (see chapter 4).

2.2.6 Expression analysis

For the expression analysis, a reference-based approach using TopHat (Trapnell et al. 2009) was carried out (Figure 2.2). Firstly, the processed RNA-Seq FASTQ reads from the three tissues, rhizines, apothecia, and thallus, collected at different times (Table 2.2) were mapped to the P.mem v1.1 (reference genome) using Bowtie2 (v2.1.0.0) (Langmead and Salzberg 2012), and TopHat (v2.0.8b) (Trapnell et al. 2009), with the default parameters. The output from TopHat, i.e., aligned reads in BAM format was viewed and analyzed using Seqmonk (<http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/v1.37.1>). Then, nine comparisons were carried out (Table 2.5).

Pseudo chromosomes were created using the build custom genome option implemented in Seqmonk and features such as mRNA, CDS, exons, and introns were added using the gff annotation file of the *P. membranacea* genome (P.mem v1.1). Datasets and replicate sets were defined accordingly using the 'edit data sets' and 'edit replicate sets', respectively, present in the data menu of Seqmonk. For the transcriptome analysis, mRNA was quantified using the RNA-Seq quantitation pipeline on merged transcripts isoforms counting reads over a mRNA feature quantified in log₂ RPM (reads per feature per million reads of (nonstrand-specific) library) and raw read counts (Wan et al. 2016). The cumulative distribution of the data was adjusted across the sample at the 75th percentile (Langmead et al. 2010). Lists of genes were generated from the normalized quantified data using an intensity difference filter (within Seqmonk) with multiple testing corrections, and an FDR (false discovery rate) correction of $p < 0.05$ (Brien et al. 2013). An intensity difference test is a pairwise statistical test used to compare two or more data. First, local distribution of the differences for probes (mRNA) with similar average intensity to the probe (mRNA) being examined is constructed. The set of

differences between the two data sets is then modeled, i.e., when the overall normal distribution of the data is examined the likelihood of whether a data is an outlier or not is evaluated. *P* value is calculated for the changes in variability at high and low intensities, and in general, highly expressed genes have low variability, and genes with low expression have high variability.

Genes are defined as differentially expressed (DE) if the fold change (FC) was greater than 2 and *P* value was lesser than 0.05. Fold Change (FC) was calculated from log₂ RPM values using the following formula: when $x > y$, then FC is calculated using $2^{(x-y)}$ and when $y > x$, FC is calculated by $0-(1/2^{(x-y)})$ (Simon Andrews, personal communication, Sep 4th, 2017). For each of the tissue samples, a pairwise comparison was made (Table 2.1). The list of genes from each comparison was collated using the 'collate existing lists' option available in the filtering menu of Seqmonk. A total of 1901 DE genes were generated from the 10 pairwise comparisons of thallus whole biological replicates (TW-B vs. TW-G vs. TW-W vs. TW-R vs. TW-Y (su-2012)) (Table 2.1b). After collating the DE genes the final list comprised of 768 genes and 720 genes that had >2X FC expression (from now on referred to as the background (BKG) genes) were removed from further tissue comparisons that contained the thallus samples (see section 2.3.5 and Table S18). The BKG genes were removed from the tissue comparisons to shed light on the other DE genes and the list of genes before and after BKG gene removal is listed in Table 2.5. Hierarchical clustering analysis and heat maps were generated for differentially expressed genes (DEGs) using CIMminer (<http://discover.nci.nih.gov/cimminer/home.do>). For the data visualization, a cumulative distribution plot, scatter plot, PCA plot, and dendrogram (data tree) were constructed using the quantitated read counts in SeqMonk.

Table 2.1 Pairwise comparisons of the three tissues collected at different time points.

(a) Thallus whole biological replicates (TW–B vs. TW–G vs. TW–W vs. TW–R vs. TW–Y (su-2012))

Pairwise comparisons	Differentially expressed genes
TW–B vs. TW–G	59
TW–B vs. TW–R	217
TW–B vs. TW–W	122
TW–B vs. TW–Y	165
TW–G vs. TW–R	233
TW–G vs. TW–W	224
TW–G vs. TW–Y	234
TW–R vs. TW–W	195
TW–R vs. TW–Y	214
TW–W vs. TW–Y	238

(b) Apothecia (au -2009) vs. Thallus Body (su-2012) vs. Rhizine (su-2012)

Pairwise comparisons	Differentially expressed genes
A vs. R	91
A vs. TB	84
R vs. TB	321

(c) Thallus whole (au -2009) vs. (au -2010) vs. (su 2011) vs. (su- 2012)

Pairwise comparisons	Differentially expressed genes
TW au-09 vs TW au-10	73
TW au-09 vs TW su-11	82
TW au-09 vs TW su-12	77
TW au-10 vs TW su-11	96
TW au-10 vs TW su-12	69
TW su-11 vs TW su-12	108

2.3 Results

2.3.1 Transcriptome sequencing and assembly

Illumina RNA–Seq reads were obtained for each of the three tissues; thalli, rhizines, and apothecia. A total of 629,789,750 Illumina reads were obtained from 13 samples. The number of raw reads obtained for each tissue after duplicate removal and the number mapped on the P.mem v1.1 genome scaffold is presented in Table 2.2.

To generate tissue-specific transcriptomes, libraries representing different tissues were *de novo* assembled with Trinity (Grabherr et al.

2011) using default parameters. In sum, 110,092 transcripts ranging from 201 bases to 18,258 bases were generated from apothecia (autumn 2009), rhizine (autumn 2010) and thallus whole (autumn 2009 and summer 2011) samples (Figure 2.3). The number of transcripts varied for each of the tissues with the highest number of transcripts for the rhizine (au-2010) (48,708) tissue followed by the thallus whole tissue (su-2011) with 32,441 transcripts, whereas the apothecia (au-2009) and thallus whole (au-2009) yielded much less, 18,551 and 10,392 transcripts, respectively (Figure 2.3).

The transcripts from all the tissues were retained and used as a redundant reference collection of ESTs/mRNAs (110,092 transcripts in total) in genome annotation to provide information on alternative splicing and gene/transcriptional unit demarcation. A total of 28,854 transcript isoforms were generated from 16,400 gene models. A summary of the transcriptome assemblies generated for each tissue is presented in Table S4.

Table 2.2 Summary of RNA–Seq experimental results.

Tissue	Season/ Year	Technical replicates	Name of the reads	No. of Raw reads	Reads after Trimming & Duplicate removal	Total no. of Mapped Reads	Mean Read Length
Apothecia	au-2009	2	A.6.1.fastq	8,255,182	5,534,033	8,223,306	34
			A.6.2.fastq	8,255,182	5,650,804		
			A.8.1.fastq	6,669,865	4,611,556		
			A8.2.fastq	6,669,865	4,741,192		
Rhizines	au-2010	1	R.1.fastq	50,695,936	24,333,745	14,932,505	65
			R.2.fastq	50,695,936	24,859,735		
Rhizine	su-2012	2	R1.1.fastq	8,969,372	5,806,099	12,423,174	86
			R1.2.fastq	8,969,372	5,880,455		
			R2.1.fastq	9,361,612	5,882,629		
			R2.2.fastq	9,361,612	6,013,842		
Thallus Body	su-2012	3	TB1.1.fastq	5,765,946	3,661,458	19,969,672	86
			TB1.2.fastq	5,765,946	3,776,073		
			TB4.1.fastq	7,378,979	4,540,825		
			TB4.2.fastq	7,378,979	4,693,124		
			TB5.1.fastq	15,974,336	9,109,141		
			TB5.2.fastq	15,974,336	9,428,133		
Thallus Whole	au-2009	2	TW.5.1.fastq	7,276,119	4,762,940	9,335,246	34
			TW.5.2.fastq	7,276,119	4,901,291		
			TW.7.1.fastq	7,125,968	4,680,033		
			TW.7.2.fastq	7,125,968	4,839,542		

Thallus Whole (T)	au-2010	1	TW.T.8.1.fastq	37,207,187	15,083,294	13,089,207	74
			TW.T.8.2.fastq	37,207,187	15,026,017		
Thallus Whole (D)	su-2011	1	TW.D.1.fastq	14,600,921	7,773,245	9,330,207	91
			TW.D.2.fastq	14,600,921	7,911,734		
Thallus Whole (W)	su-2011	1	TW.W.1.fastq	16,560,736	8,961,224	10,564,774	91
			TW.W.1.fastq	16,560,736	8,801,929		
Thallus Whole (TW _B)	su-2012	2	TWB2.1.fastq	3,558,803	2,385,051	3,758,603	85
			TWB2.2.fastq	3,558,803	2,422,337		
			TWB3.1.fastq	1,305,769	956,624		
			TWB3.2.fastq	1,305,769	970,738		
Thallus Whole (TW _G)	su-2012	3	TWG1.1.fastq	10,033,987	5,766,883	28,275,595	85
			TWG1.2.fastq	10,033,987	5,785,200		
			TWG4.1.fastq	16,936,084	9,248,448		
			TWG4.2.fastq	16,936,084	9,241,422		
			TWG5.1.fastq	16,080,188	8,778,978		
			TWG5.2.fastq	10,080,188	8,802,298		
Thallus Whole (TW _R)	su-2012	2	TWR2.1.fastq	21,127,889	10,804,912	13,865,600	85
			TWR2.2.fastq	21,127,889	10,807,766		
			TWR3.1.fastq	3,523,575	2,177,399		
			TWR3.2.fastq	3,523,575	2,217,911		
Thallus Whole (TW _W)	su-2012	2	TWW1.1.fastq	8,554,801	5,006,784	15,508,846	84
			TWW1.2.fastq	8,554,801	5,020,339		
			TWW2.1.fastq	16,475,430	8,499,437		
			TWW2.2.fastq	16,475,446	8,983,995		
Thallus Whole (TW _Y)	su-2012	2	TWY1.1.fastq	16,115,479	8,482,539	15,399,551	84
			TWY1.2.fastq	16,115,479	8,490,683		
			TWY2.1.fastq	8,340,703	4,670,581		
			TWY2.2.fastq	8,340,703	4,683,986		

The biological replicates are represented in alphabetical subscripts.

Transcript length distribution

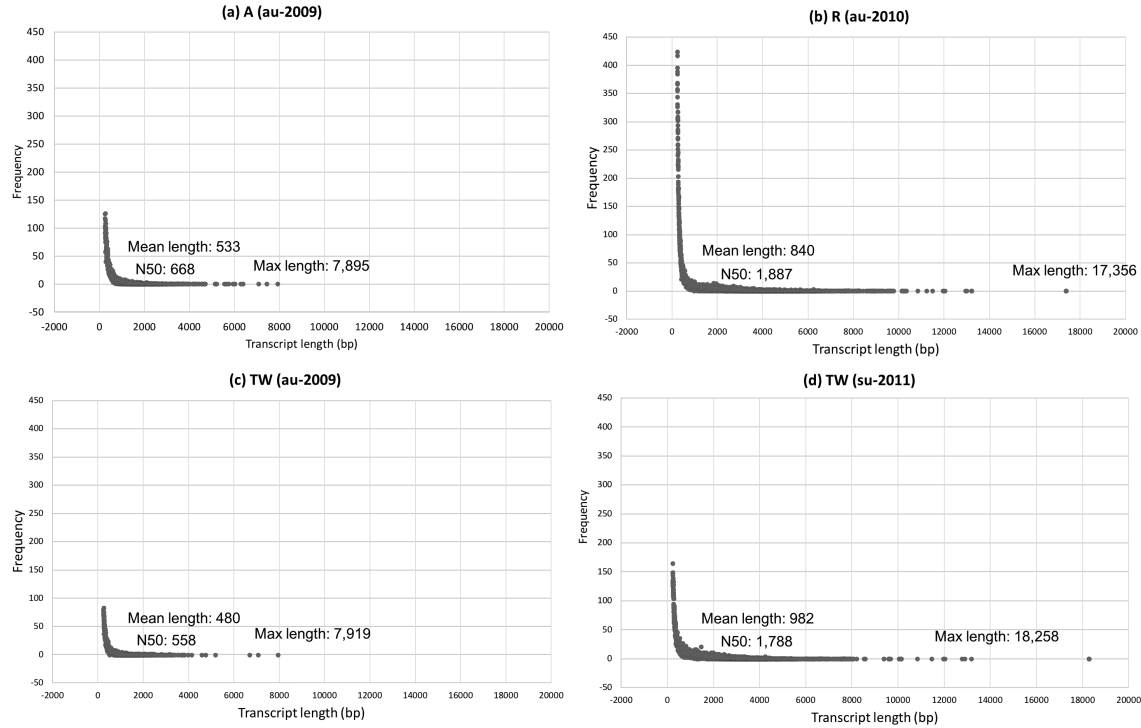


Figure 2.3 Length distribution of *de novo* assembled transcripts.

The total number of transcripts for each of the datasets. (a) Apothecia (au-2009). (b) Rhizine (au-2010). (c) Thallus whole (au-2009). (d) Thallus whole (su-2011).

2.3.2 BUSCO analysis

The genome completeness of the *P. membranacea* genome (P.mem v1.1) and transcriptome (Trinity *de novo* transcriptomes generated from the three tissues: Apothecia (au-2009), Rhizines (au-2010), and Thallus whole (au-2009 and su-2011) assemblies were assessed by CEGMA (Core Eukaryotic Genes Mapping Approach) (Parra et al. 2007) and BUSCO v3 (Benchmarking Universal Single-Copy Orthologs) (Simao et al. 2015). CEGMA estimates the completeness by searching for a set of 248 general single-copy orthologs (Parra et al., 2007), whereas the BUSCO genes are based on evolutionarily-informed expectations of gene content from near-universal single-copy orthologs from OrthoDBV9 (<https://www.orthodb.org/>) (Simao et al. 2015). For fungi, the BUSCO reference dataset comprises 1438 genes. The results of BUSCO analyses are presented in Figure 2.4, and additional files are summarized in Appendix C.

Genes that are identified from BUSCO analysis are classified as either complete, fragmented, duplicated and missing. Analysis of the P.mem v1.1 genome showed that 90% and 6.8% of the 1438 expected fungal genes were complete and fragmented, respectively, with an additional 0.69% duplicated, and 3.05% missing (Figure 2.4, Table S5). By this measure, the *P. membranacea* genome assembly (P.mem v1.1) represents 96.9% of the *P. membranacea* genome. Forty-four, 360, 19, 846, and 31 BUSCOs were missing from the *P. membranacea* P.mem v1.1, apothecia (au-2009), rhizine (au-2010), thallus whole (au-2009) and thallus whole (su-2011) assemblies, respectively (Figure 2.4 and Table S5). The 44 missing genes from the *P. membranacea* genome were further investigated. A local TBLASTN search (Altschul et al. 1990) was performed against the P.mem v1.1 genome assembly. Out of 44 missing BUSCO sequences, 38 were represented in a *P. membranacea* rhizine transcriptome assembly as complete (28), duplicated (6), and fragmented (4). Six sequences were absent in both genome and transcriptome assemblies, of which one BUSCO (EOG7ZPNX4) was absent in the *P. membranacea* genome and transcriptome assemblies, as well as the reference genome, *A. nidulans* (Table S5). The missing BUSCO proteins included ribosomal proteins and sequences that contain repeated sequences that impair the assembly process. Mis-assembly can also lead to fragmentation of genes into different scaffolds. However, sequences with significant homology to the missing BUSCOs were present in the *P. membranacea* genome and transcriptome, thus indicating that the genome assembly of *P. membranacea* v1.1 is ~99.5% complete with regard to

protein coding genes. A well assembled *Aspergillus nidulans* genome (Cerqueira et al. 2014) in comparison contained 99.6% of complete BUSCOs and five BUSCOs missing (Figure 2.4 and Table S5). The complete and fragmented BUSCO orthologs ranged between 99.1% and 99.7% for other fungal genomes such as *A. nidulans*, *M. brunnea*, *B. cinerea*, *S. sclerotiorum*, and *M. oryzae* (Figure 2.4 and Table S5) (Neu et al. 2017).

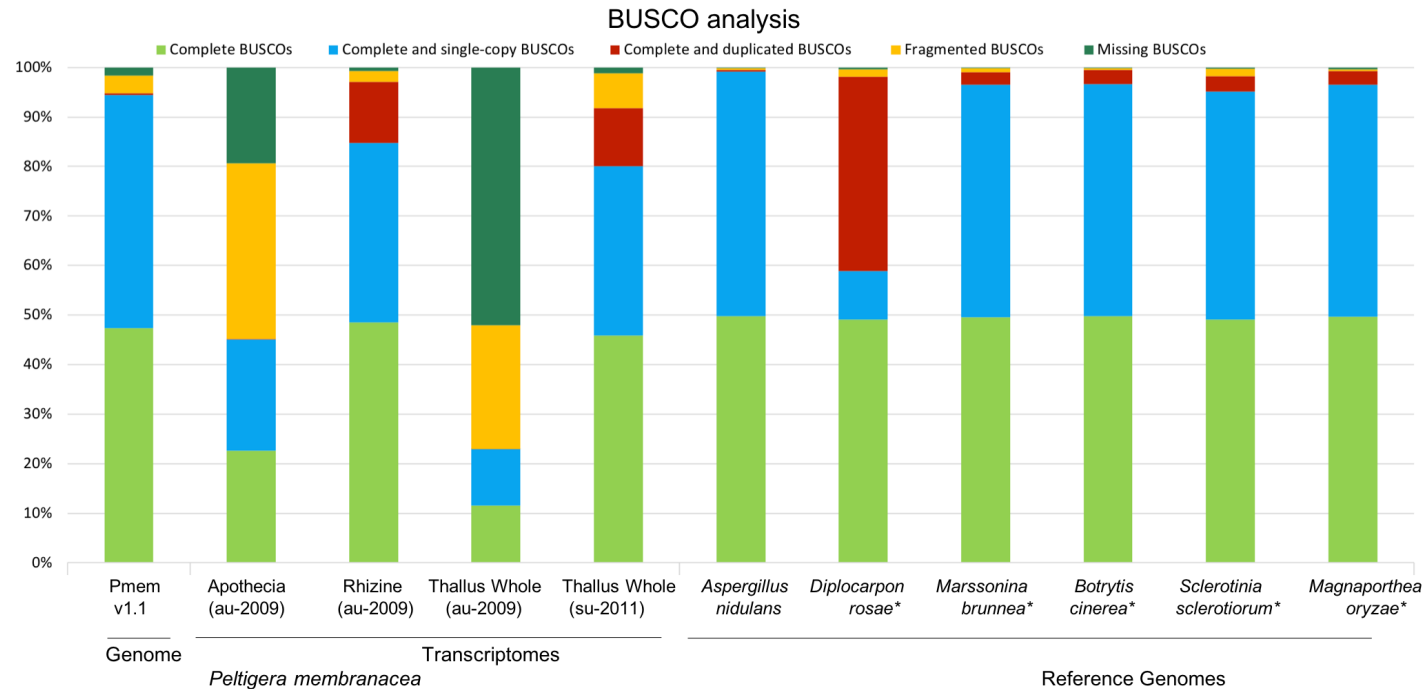


Figure 2.4 BUSCO analysis.

* Reference genomes from GenBank assembly accession nos. GCA_002317995.1 (*D. rosae*), GCA_000298775.1 (*M. brunnea*), GCA_000143535.4 (*B. cinerea*), GCA_000146945.2 (*S. sclerotiorum*) and GCA_000002495.2 (*M. oryzae*).

2.3.3 Data quality control and gene prediction

The measurement of transcript abundance can be biased due to lower recovery of longer transcripts, RNA fragmentation, cDNA synthesis artifacts where adapters and barcodes during ligation influence cDNA synthesis efficiency, GC content where local RNA G/C-content influences PCR reactions as well as differences in poly-A selection and mRNA degradation (Finotello et al. 2014). The heat map generated for 3' bias showed no obvious biases within the RNA samples indicating a low level of degradation (Figure S6). Since the biases were minimal between the samples, the measured raw counts were not corrected.

A total of 16,400 gene models with open reading frames longer than 200 bases with 28,854 isoforms were predicted and annotated. The transcript number for each annotated gene model varied. Out of the 16,400 gene models, 10,310 (62.9%) were represented by one isoform, 3542 (21.6%) by two isoforms, and 2548 (15.5%) of the gene models had >2 isoforms. The highest number of isoforms for a gene model was 20. For further downstream analysis, the proteins derived from the 16,400 gene models were subjected to Blast2GO annotation. Categorized by size 491 were ≥ 1000 amino acids (aa) long, 2380 were ≥ 500 to ≤ 999 aa long, 9513 were ≥ 100 to ≤ 499 aa long, and 4016 were ≥ 67 to ≤ 99 aa long.

The *P. membranacea* (P.mem v1.1) genome assembly has a size of ~ 43 Mbp that consists of ~ 1 Mbp in repeats (2.44%) (Table S29). The CDS and intron regions were extracted from the *P. membranacea* genome using the 'extract annotations' available in CLC Genomics Workbench v9.5.1. Using an in-house Perl script, the number of bases in the CDS and introns were deduced. The size of CDS and introns are ~ 14.6 Mbp and ~ 3 Mbp, respectively. The remaining ~ 24.4 Mbp may comprise UTRs (5' and 3'), rDNAs, tRNAs, miRNAs, lncRNAs, centromeres, telomeres and other features.

2.3.3.1 Comparison of the *de novo* transcriptome assembly and reference genome annotation (P.mem v1.1)

A total of 2,068 *de novo* transcripts did not map to the annotated genes. The unmapped *de novo* transcripts were used as queries in a BLAST search against the reference genome (P.mem v1.1). 1507 transcripts were mapped, and 561 transcripts were unmapped (Figure 2.5, Table S11). BLASTX analysis of the 561 unmapped transcripts showed that 201 transcripts had a significant BLAST hit, with 58 transcripts being hypothetical proteins. There were proteins containing repeat domains, RNA directed RNA polymerase, ribosomal proteins, and polyproteins. The BLAST results of all the unmapped transcripts are presented in Table S12 to S15.

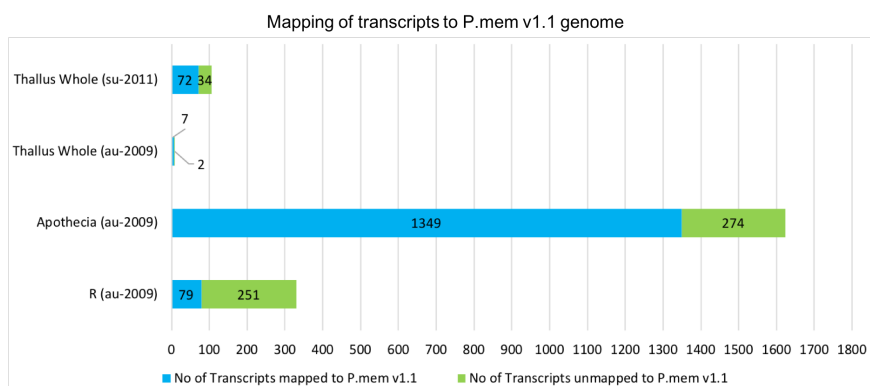


Figure 2.5 Number of missing and unmapped *de novo* transcripts from P.mem v1.1 genome.

2.3.3.2 Validation of predicted genes using curated genes

To assess the quality of predicted gene models, proteins derived from the 16,400 gene models were aligned with few manually curated proteins available in NCBI. The list of proteins included housekeeping genes, a high-affinity ammonium transporter, mating type switching protein, polyketide synthase genes, and other ascomycete proteins (Table 2.3). Overall, satisfactory performance of the *de novo* transcriptome-based predictions using the AUGUSTUS tool was indicated (Figure 2.6 and Appendix F).

Full-length alignments with 100% identities were produced for the following manually curated *P. membranacea* proteins: high-affinity ammonium transporter, LEC-2, elongation factor-1 alpha, DNA directed RNA polymerase II, and beta-tubulin (Figure 2.6 and Figures S7–S20). However, manually curated *P. membranacea* proteins such as TBP (TATA-binding protein) associated factor, heat shock protein (HSP98), and Peptidase A22B were either missing amino or carboxy terminal amino acids (Table 2.3). Therefore, for the above genes, protein homologues from other ascomycetes were included in the alignment, and better alignments were produced (Appendix F). An example of beta-tubulin, PP1 and PKS6 are shown in Figure 2.6A. Except for PKS6, no full-length alignments were produced for PKSs, e.g. PKS1 and 7 were aligned to three gene models with different transcripts (Figure 2.7). PKS9 was absent from P.mem v1.1 and present only in 454 contigs (Pmem 454Allcontigs) (Table. 2.3).

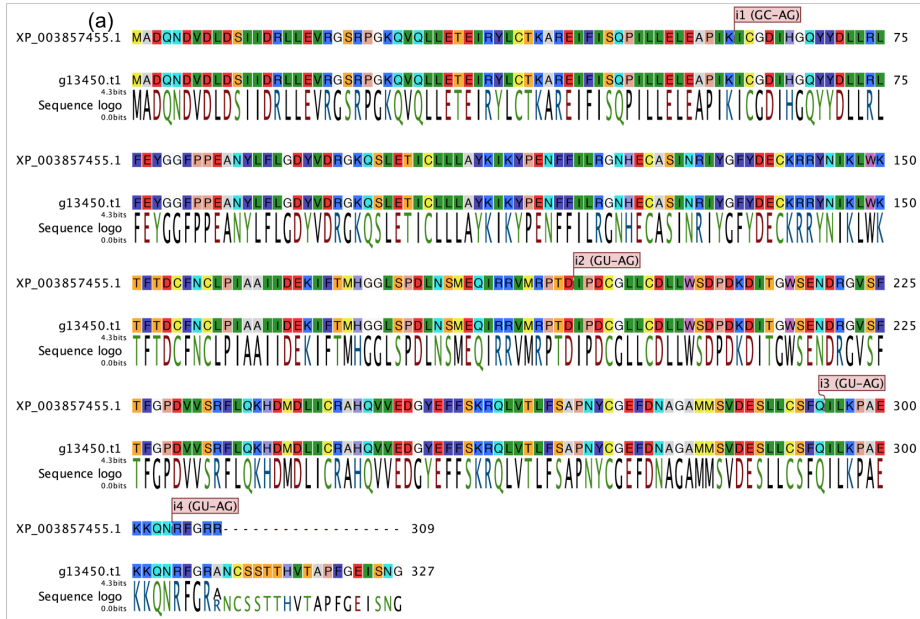
Standard GU-AG splice sites agreed with gene curation of spliced transcripts. GC-AG splice sites that are known to exist in fungi were also deduced accurately (Figure 2.6 and Table 2.3). Intron 1 (GC-AG) of *pp1* (Figure 2.6b) and intron 3 (GC-AG) of the gene encoding a mitochondrial carrier protein (Figure 2.6c) were predicted accurately. Curated proteins such as SWI-10 and PP1 were available only partially and using the redundant reference transcriptome of *P. membranacea* the full-length proteins were predicted for the above two proteins (Figure S15, Figure 2.6b). Five transcripts with varying lengths were generated from the RNA-Seq data available for the *lec-1* gene, and they differed from the curated version at the 5' end (Figure S10). The 5' end of *lec-1* has now been added to the sequence (accession no: JQ899138.1) originally submitted to GenBank.

Table 2.3 List of curated proteins used in the gene prediction validation study.

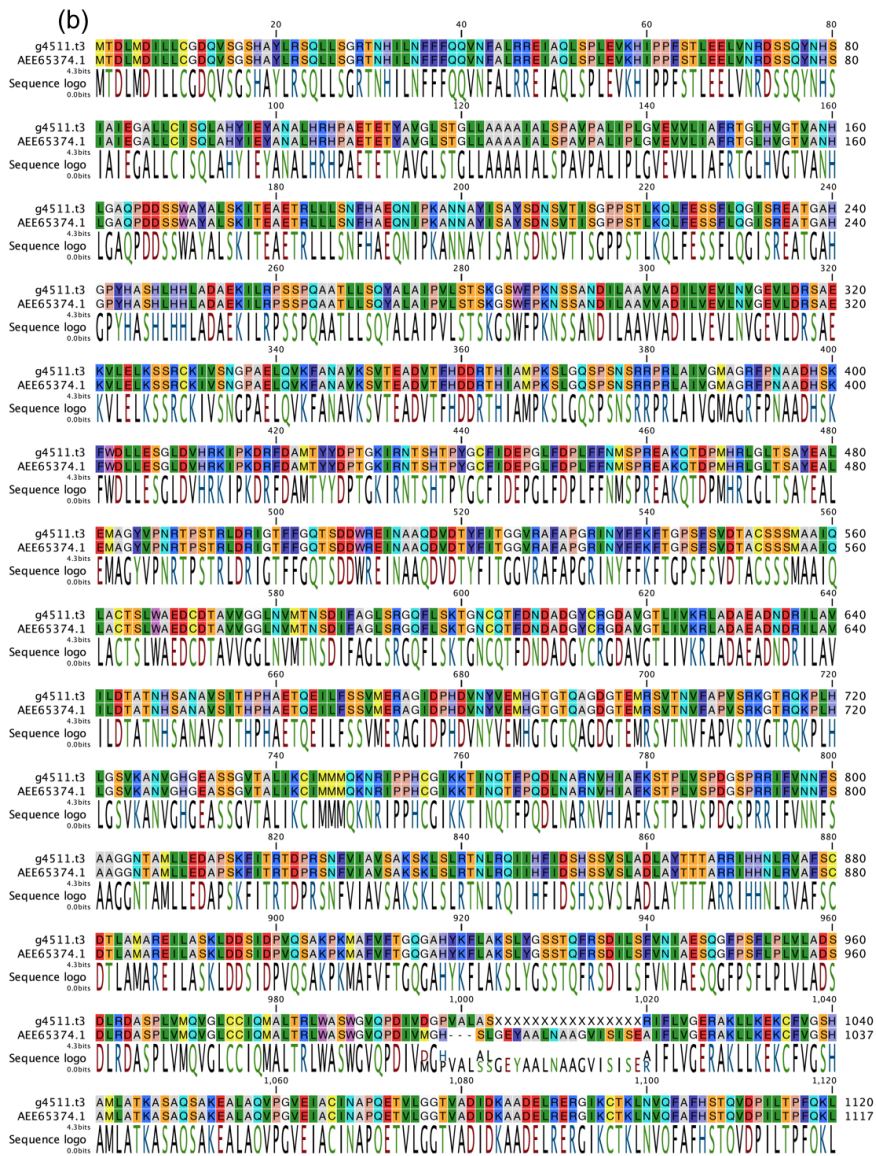
S. NO	Accession	Description	Gene name	Organism	<i>P. membranacea</i> gene id.	Scaffold Name	No. of Introns	Acceptor-Donor site	Identities	Gaps	References
1	AGM18788.1	High affinity ammonium transporter	<i>amt</i>	<i>Peltigera membranacea</i>	g16330.t1	scaff35579	2	11: GU-AG 12: GU-AG	152/152 (100%)	0/152 (0%)	McDonald et al. (2013)
2	AGC12988.1	Galectin like protein -1	<i>lec-1</i>	<i>Peltigera membranacea</i>	g11004.t1, g11004.t2, g11004.t3, g11004.t4, g11004.t5	scaff31342		11: GU-AG	168/168 (100%)	0/168 (0%)	Miao et al. (2012)
3	AGC54757.1	Galectin like protein -2	<i>lec-2</i>	<i>Peltigera membranacea</i>	g15502.t1	scaff49756	1	11: GU-AG	157/159 (98%)	0/159 (0%)	Manoharan et al. (2012)
4	ACU09201.1	RNA polymerase II large subunit, partial	<i>rpb1</i>	<i>Peltigera membranacea</i>	g1188.t1, g1188.t2, g1188.t3	scaff33055	2	11: GU-AG 12: GU-AG	184/184 (100%)	0/184 (0%)	O'Brien et al. (2009)
5	AFQ55277.1	Elongation factor-1 alpha	<i>ef-1</i>	<i>Peltigera membranacea</i>	g5395.t1	scaff31241	5	11: GU-AG 12: GU-AG 13: GU-AG 14: GU-AG 15: GU-AG	459/459 (100%)	0/459 (0%)	Xavier et al. (2012)
6	AQM49832.1 XP_002791902.1	Peptidase A22B		<i>Peltigera membranacea</i> <i>Paracoccidioides lutzii</i> Pb01	g9385.t1	scaff31558	0		527/527 (100%) 231/487 (47%)	0/527 (0%) 56/487 (11%)	Desjardins et al. (2011)
7	AQM49833.1	mat-ing-type switching protein, partial	<i>swi10</i>	<i>Peltigera membranacea</i>	g15111.t1	scaff52844	0		321/321 (100%)	0/321 (0%)	Steinhauser et al. (2016)
8	AQM49836.1 GAA83384	TATA binding protein associated factor	<i>tbp</i>	<i>Peltigera membranacea</i> <i>Aspergillus kawachii</i> IFO 4308	g1061.t1	scaff30505	3	11: GU-AG 12: GU-AG 13: GU-AG	1218/1887 (64%)	34/1887 (2%)	Steinhauser et al. (2016) Futagami et al. (2011)
9	AFQ55279.1	DNA directed RNA polymerase II		<i>Peltigera membranacea</i>	g3092.t1	scaff31547	1	11: GU-AG	1255/1255 (100%)	0/1255 (0%)	Xavier et al. (2012)
10	AQM49829.1 XP_002796242	heat shock protein	<i>hsp98</i>	<i>Peltigera membranacea</i> <i>Paracoccidioides lutzii</i> Pb01	g10866.t1	scaff45564	0		606/709 (85%)	2/709 (0%)	Steinhauser et al. (2016) Desjardins et al. (2011)
11	AFJ45056.1	Beta tubulin	<i>tub-2</i>	<i>Peltigera membranacea</i>	g13648.t1	scaff34161	8	11: GU-AG 12: GU-AG 13: GU-AG 14: GU-AG 15: GU-AG 16: GU-AG 17: GU-AG 18: GU-AG	426/426 (100%)	0/426 (0%)	Miao et al. (2012)
12	ADF28668.2	Reducing type I polyketide synthase	<i>pk1</i>	<i>Peltigera membranacea</i>	g10700.t1, g10701.t1, g10701.t2, g10701.t3, g10701.t4, g10701.t5, g10701.t6, g10701.t7, g10702.t1	scaff37726	—	—	—	—	Kampa et al. (2013)
13	ADF28669.1	Reducing type I polyketide synthase	<i>pk2</i>	<i>Peltigera membranacea</i>	g10421.t1, g10421.t2, g10421.t3, g10421.t4, g10421.t5, g10421.t6, g10421.t7, g10421.t8, g10421.t9, g10421.t10, g10421.t11, g10421.t12, g10421.t13, g10421.t14, g10421.t15	scaff32161	—	—	—	—	Kampa et al. (2013)
14	ADF28670.1	Reducing type I polyketide synthase	<i>pk3</i>	<i>Peltigera membranacea</i>	g8391.t3	scaff35874 (scaff32161)	—	—	—	—	Kampa et al. (2013)
15	AEE65372.1	Non reducing type I polyketide synthase	<i>pk4</i>	<i>Peltigera membranacea</i>	g10571.t1, g10571.t2, g10571.t3	scaff30968	—	—	—	—	Kampa et al. (2013)
16	AEE65373.1	Reducing type I polyketide synthase	<i>pk5</i>	<i>Peltigera membranacea</i>	g8166.t1, g8167.t1, g8168.t1, g8168.t2,	scaff42419	—	—	—	—	Kampa et al. (2013)

17	AEE65374.1	Non reducing type I polyketide synthase	<i>pks6</i>	<i>Peltigera membranacea</i>	g4511.t3	scaff38861	-	-	2104/2128 (98%)	3/2128 (0%)	Kampa et al. (2013)
18	AEE65375.1	Reducing type I polyketide synthase	<i>pks7</i>	<i>Peltigera membranacea</i>	g6020.t1, g6020.t2, g6020.t3, g6020.t4, g6021.t1, g6022.t1	scaff37151	-	-	-	-	Kampa et al. (2013)
19	AEE65376.1	Reducing type I polyketide synthase	<i>pks8</i>	<i>Peltigera membranacea</i>	-	scaff51530	-	-	-	-	Kampa et al. (2013)
20	AEE65377.1	Reducing type I polyketide synthase	<i>pks9</i>	<i>Peltigera membranacea</i>	-	contig05313 contig12953 contig50194 contig32425	-	-	-	-	Kampa et al. (2013)
21	AEE87273.1	Reducing type I polyketide synthase	<i>pks10</i>	<i>Peltigera membranacea</i>	g10306.t1, g10306.t2, g10306.t3, g10306.t4, g10306.t5, g10306.t6, g10306.t7, g10306.t8, g10306.t9, g10306.t10, g10306.t11, g10306.t12, g10306.t13	scaff30823	-	-	-	-	Kampa et al. (2013)
22	AEE87274.1	Reducing type I polyketide synthase	<i>pk11</i>	<i>Peltigera membranacea</i>	g10449.t1, g10449.t2, g10449.t3, g10449.t4, g10449.t5, g10449.t6, g10449.t7, g10449.t8, g10449.t9	scaff39401	-	-	-	-	Kampa et al. (2013)
23	ODM17725	Glucose-6-phosphate isomerase	<i>gpi</i>	<i>Aspergillus cristatus</i>	g4152.t1	scaff32335	2	i1: GU-AG i2: GU-AG	484/552 (87%)	0/552 (0%)	Ge et al. (2016)
24	XP_003719976	Aldehyde dehydrogenase	<i>adh</i>	<i>Magnaporthe oryzae 70-15</i>	g11180.t1	scaff30757	3	i1: GU-AG i2: GU-AG i3: GU-AG	411/496 (82%)	0/496 (0%)	Dean et al. (2005)
25	XP_003857455.1	pp1, catalytic subunit of type 1 serine/threonine protein phosphatase	<i>pp1</i>	<i>Zymoseptoria tritici</i>	g13450.t1	scaff36706	4	i1: GC-AG i2: GU-AG i3: GU-AG i4: GU-AG	308/308 (100%)	0/308 (0%)	Goodwin et al. (2011)
26	XP_002790124	mitochondrial carrier protein		<i>Paracoccidioides lutzii Pb01</i>	g10596.t1	scaff43257	6	i1: GU-AG i2: GU-AG i3: GC-AG i4: GU-AG i5: GU-AG i6: GU-AG	205/255 (80%)	1/255 (0%)	Desjardins et al. (2011)

Sequence alignment of PP1 proteins



Sequence alignment of PKS6 proteins



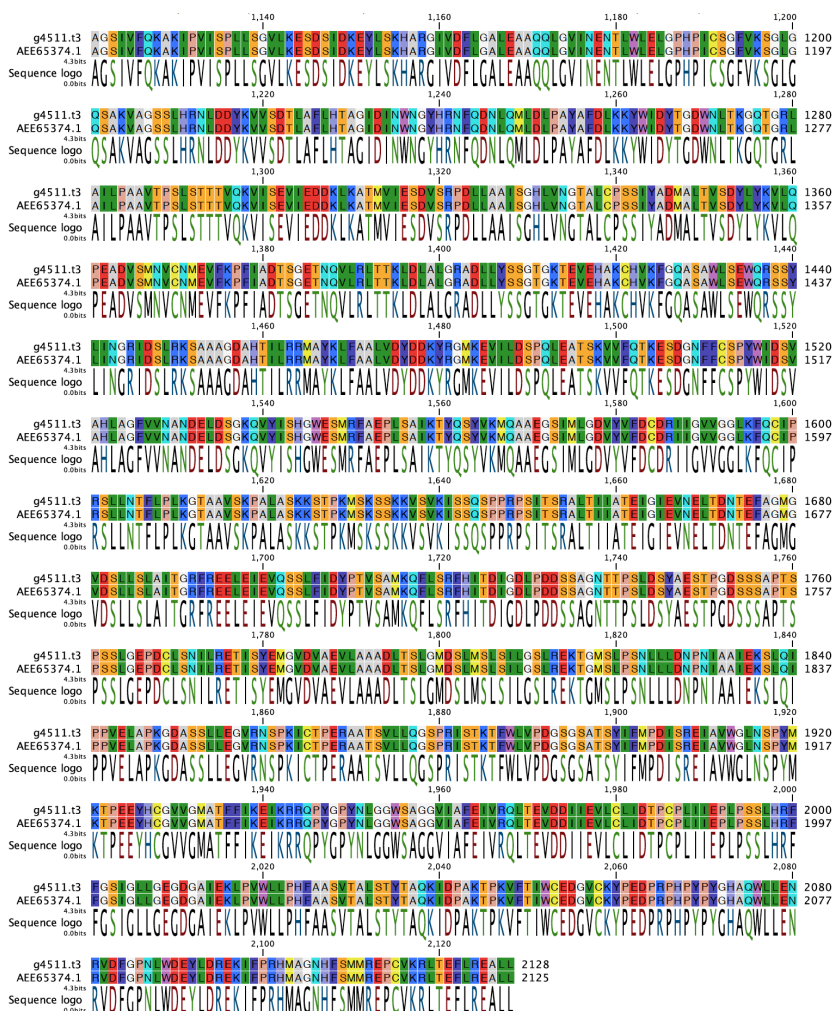


Figure 2.6 Alignment of predicted and curated proteins. Ungapped alignment of *Zymoseptoria tritici* curated PP1 (accession no. XP_003857455.1) and predicted PP1 g13450.t1. The non-canonical splice sites of PP1 are indicated. (b) Alignment of *P. membranacea* curated PKS6 (accession no. AEE65374.1) and predicted PKS6 g4511.t3.



Figure 2.7 Mapping of PKS genes to P.mem v1.1 genome.

(a) ADF28668.2 (PKS1) (b) AEE65374.1 (PKS6) (c) AEE65375.1 (PKS7) on *P. membranacea* annotated genes (P.mem v1.1). Only one representative of the transcript for a gene is depicted.

2.3.4 Functional annotation and metabolic pathways

AUGUSTUS derived proteins from *P. membranacea* were searched against the NCBI nr (non-redundant) database. Around 82.2% had a BLAST hit ($p < 1 \times 10^{-6}$), inclusive of proteins with or without functional annotation, and 38% of ORFs had only a BLAST hit with no functional information (Figure 2.8). The remaining protein models (18%, 2927 proteins) had no known function (neither BLAST hit nor functional information) in any database but were expressed in one or all of the tissues. The Blast2GO summary is presented in Figure 2.8

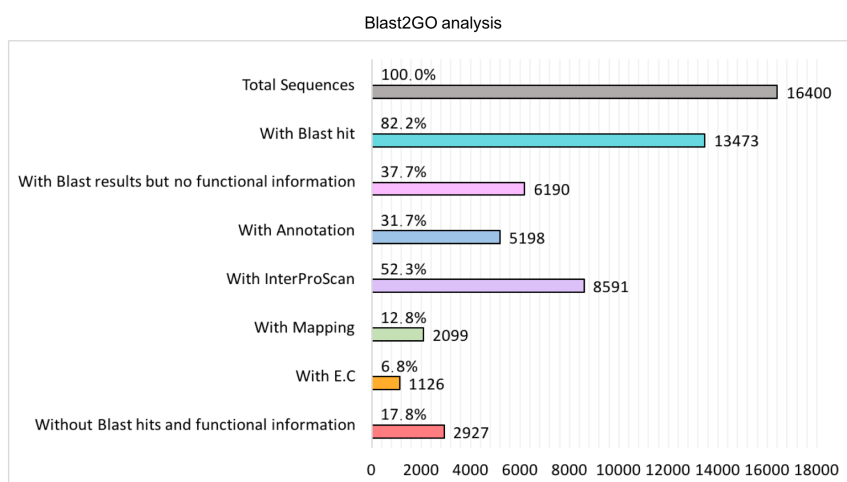


Figure 2.8 Summary of Blast2GO analysis.

For GO (gene ontology) analysis, a total of 19,233 GO terms were assigned to 7,282 proteins (44.4%). In the GO classification, proteins are classified into three core categories: biological process (BP), cellular component (CC) and molecular function (MF). Most of the GO terms were assigned to biological processes (7,911; 41.1%), followed by cellular components (6,120; 31.8%), and molecular functions (5,202; 27%). The terms included 20 different functional groups at hierarchy level 2. Within the biological processes, the top three assigned processes were metabolic processes (32%), cellular processes (26%) and single organism processes (17%). In the cellular component processes, the majority were assigned to cell (21.7%), cell part (21.6%) and organelle (14.8%), and within molecular functions, the majority were attributed to binding (55.3%) and catalytic activity (44.6%) (Figure 2.9). All of the GO

terms were inferred from the UniProtKB database (UniProt 2010). The number of protein sequences assigned to each category is shown as pie charts in Figure 2.9.

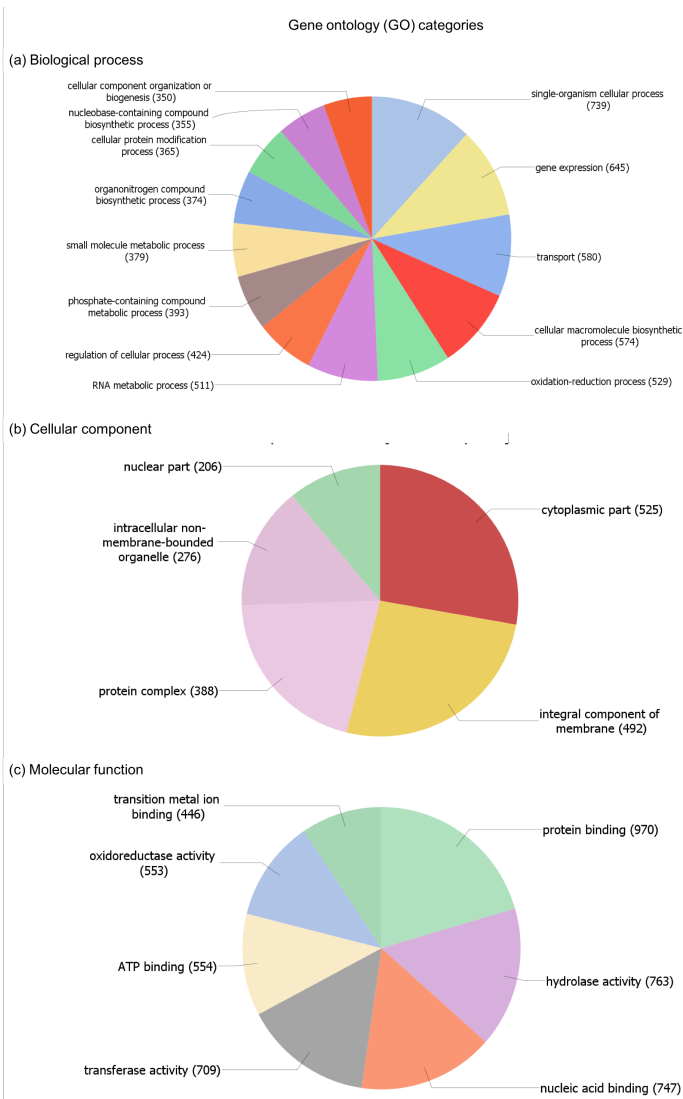


Figure 2.9 Distribution of P.mem v1.1 sequences by Gene Ontology (GO) at hierarchy level 2.

Pie charts of GO terms for the three categories (a) Biological Process (BP) (b) Cellular Component (CC) and (c) Molecular Function (MF).

Following annotation with Blast2GO, K numbers (KEGG Orthology identifiers) were assigned by submitting 16,400 protein sequences submitted to BlastKOALA and GhostKOALA. A total of 3616 K numbers were assigned. The K numbers were mapped to a total of 137 KEGG metabolic pathways (Table 2.4, Figure 2.9b and 2.10a). Thirty-four pathways, including the key central metabolic pathways involving carbon, fatty acid, nitrogen, sulfur, pyruvate, purine, pyrimidine metabolism were complete (see Table 2.4, for the complete list of pathways). The most highly represented pathways were biosynthesis of secondary metabolites (232) (Figure S24), biosynthesis of amino acids (88) (Figure S25), purine metabolism (79) (Figure S26), carbon metabolism (71) (Figure S27), and oxidative phosphorylation (70) (Figure S28). Twenty-five pathways were removed because they were either represented by one enzyme or were incorrectly assigned (Table S16). For e.g., *P. membranacea* assignments to the streptomycin biosynthesis pathway (Figure S29) consist of only four KO entries: K00844 (hexokinase), K01092 (myo-inositol-1(or 4)-mono phosphatase), K01835 (phosphoglucomutase), and K01858 (myo-inositol-1-phosphate synthase). K01092 can also be involved in inositol phosphate metabolism and phosphatidylinositol signaling whereas K00844 and K01835 can be involved in 17 (http://www.genome.jp/dbget-bin/www_bget?K00844) and 11 (http://www.genome.jp/dbget-bin/www_bget?K01835) pathways, respectively. This brings the number of putative KEGG metabolic pathways in *P. membranacea* down to 112.

The critical role of gene expression in different pathways involved in carbon and nitrogen metabolism (see section 2.3.6), and melanin biosynthesis (see section 2.3.8) are examined in detail. Additionally, the glyoxylate cycle (see section 2.3.9) was identified in *P. membranacea*. A summary of all the pathways is listed in Table S17.

Table 2.4 List of the metabolic pathways identified in *P. membranacea*.

Pathway ID and Pathway	# Enzs identified
Metabolism	
Global and overview maps	
01100 Metabolic pathways	593
01110 Biosynthesis of secondary metabolites	232
01120 Microbial metabolism in diverse environments	119
01130 Biosynthesis of antibiotics	172
01200 Carbon metabolism	71
01210 2-Oxocarboxylic acid metabolism	27
01212 Fatty acid metabolism	19
01230 Biosynthesis of amino acids	88
01220 Degradation of aromatic compounds	3
Carbohydrate metabolism	
00010 Glycolysis / Gluconeogenesis	25
00020 Citrate cycle TCA cycle	21
00030 Pentose phosphate pathway	17
00040 Pentose and glucuronate interconversions	6
00051 Fructose and mannose metabolism	15
00052 Galactose metabolism	13
00053 Ascorbate and aldarate metabolism	2
00500 Starch and sucrose metabolism	18
00520 Amino sugar and nucleotide sugar metabolism	23
00620 Pyruvate metabolism	23
00630 Glyoxylate and dicarboxylate metabolism	16
00640 Propanoate metabolism	16
00650 Butanoate metabolism	12
00660 C5-Branched dibasic acid metabolism	3
00562 Inositol phosphate metabolism	17
Energy metabolism	
00190 Oxidative phosphorylation	70
00195 Photosynthesis	6
00196 Photosynthesis - antenna proteins	5
00710 Carbon fixation in photosynthetic organisms	15
00720 Carbon fixation pathways in prokaryotes	8
00680 Methane metabolism	14
00910 Nitrogen metabolism	10
00920 Sulfur metabolism	12
Lipid metabolism	
00061 Fatty acid biosynthesis	7
00062 Fatty acid elongation	7
00071 Fatty acid degradation	8
00072 Synthesis and degradation of ketone bodies	3
00100 Steroid biosynthesis	15

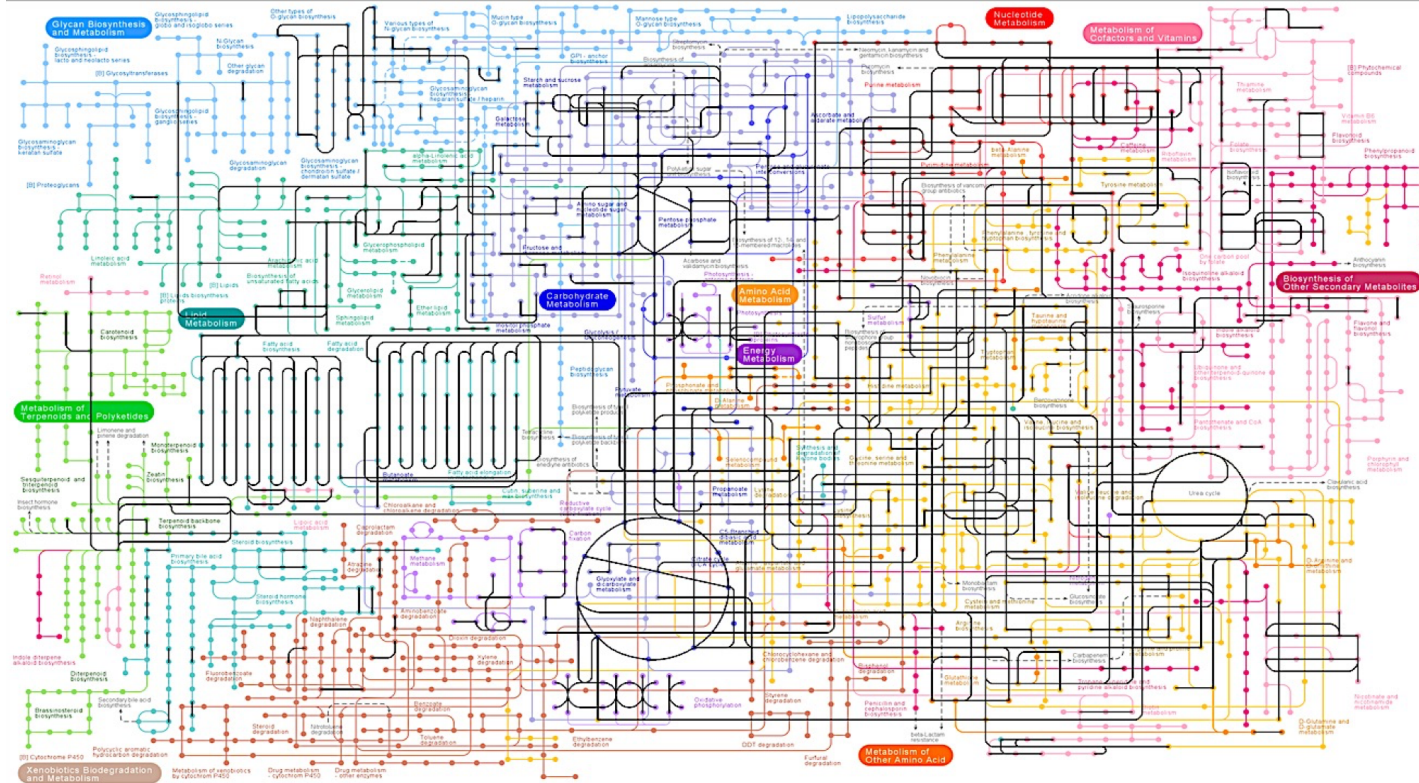
00120 Primary bile acid biosynthesis	2
00140 Steroid hormone biosynthesis	4
00561 Glycerolipid metabolism	18
00564 Glycerophospholipid metabolism	32
00565 Ether lipid metabolism	8
00600 Sphingolipid metabolism	12
00590 Arachidonic acid metabolism	4
00591 Linoleic acid metabolism	3
00592 alpha-Linolenic acid metabolism	3
01040 Biosynthesis of unsaturated fatty acids	8
Nucleotide metabolism	
00230 Purine metabolism	79
00240 Pyrimidine metabolism	61
Amino acid metabolism	
00250 Alanine, aspartate and glutamate metabolism	23
00260 Glycine, serine and threonine metabolism	30
00270 Cysteine and methionine metabolism	30
00280 Valine, leucine and isoleucine degradation	22
00290 Valine, leucine and isoleucine biosynthesis	10
00300 Lysine biosynthesis	11
00310 Lysine degradation	18
00220 Arginine biosynthesis	15
00330 Arginine and proline metabolism	17
00340 Histidine metabolism	12
00350 Tyrosine metabolism	13
00360 Phenylalanine metabolism	11
00380 Tryptophan metabolism	13
00400 Phenylalanine, tyrosine and tryptophan biosynthesis	16
Metabolism of other amino acids	
00410 beta-Alanine metabolism	12
00430 Taurine and hypotaurine metabolism	4
00440 Phosphonate and phosphinate metabolism	4
00450 Selenocompound metabolism	7
00460 Cyanoamino acid metabolism	3
00480 Glutathione metabolism	17
Glycan biosynthesis and metabolism	
00510 N-Glycan biosynthesis	27
00513 Various types of N-glycan biosynthesis	22
00515 Mannose type O-glycan biosynthesis	1
00514 Other types of O-glycan biosynthesis	2
00533 Glycosaminoglycan biosynthesis - keratan sulfate	1
00531 Glycosaminoglycan degradation	2
00563 Glycosylphosphatidylinositol GPI-anchor biosynthesis	16
00603 Glycosphingolipid biosynthesis - globo and isoglobo series	1

00604 Glycosphingolipid biosynthesis - ganglio series	1
00511 Other glycan degradation	4
Metabolism of cofactors and vitamins	
00730 Thiamine metabolism	7
00740 Riboflavin metabolism	9
00750 Vitamin B6 metabolism	6
00760 Nicotinate and nicotinamide metabolism	13
00770 Pantothenate and CoA biosynthesis	13
00780 Biotin metabolism	3
00785 Lipoic acid metabolism	3
00790 Folate biosynthesis	14
00670 One carbon pool by folate	11
00830 Retinol metabolism	2
00860 Porphyrin and chlorophyll metabolism	15
00130 Ubiquinone and other terpenoid-quinone biosynthesis	8
Metabolism of terpenoids and polyketides	
00900 Terpenoid backbone biosynthesis	18
00909 Sesquiterpenoid and triterpenoid biosynthesis	4
00906 Carotenoid biosynthesis	3
00981 Insect hormone biosynthesis	1
00908 Zeatin biosynthesis	1
00903 Limonene and pinene degradation	1
00281 Geraniol degradation	1
01051 Biosynthesis of ansamycins	1
Biosynthesis of other secondary metabolites	
00940 Phenylpropanoid biosynthesis	2
00901 Indole alkaloid biosynthesis	1
00950 Isoquinoline alkaloid biosynthesis	6
00960 Tropane, piperidine and pyridine alkaloid biosynthesis	5
00232 Caffeine metabolism	2
00965 Betalain biosynthesis	3
00332 Carbapenem biosynthesis	2
00261 Monobactam biosynthesis	3
00521 Streptomycin biosynthesis	4
00524 Neomycin, kanamycin and gentamicin biosynthesis	1
00401 Novobiocin biosynthesis	2
00405 Phenazine biosynthesis	1
00333 Prodigiosin biosynthesis	2
00254 Aflatoxin biosynthesis	1
Xenobiotics biodegradation and metabolism	
00362 Benzoate degradation	3
00627 Aminobenzoate degradation	5
00364 Fluorobenzoate degradation	1
00625 Chloroalkane and chloroalkene degradation	3

00361 Chlorocyclohexane and chlorobenzene degradation	2
00623 Toluene degradation	2
00643 Styrene degradation	1
00791 Atrazine degradation	2
00930 Caprolactam degradation	2
00621 Dioxin degradation	1
00626 Naphthalene degradation	2
00624 Polycyclic aromatic hydrocarbon degradation	1
00980 Metabolism of xenobiotics by cytochrome P450	4
00982 Drug metabolism - cytochrome P450	4
00983 Drug metabolism - other enzymes	6

Complete pathways are highlighted in red font. Twenty-five pathways were removed due to improper assignments.

KEGG metabolic pathways



KEGG metabolic pathways

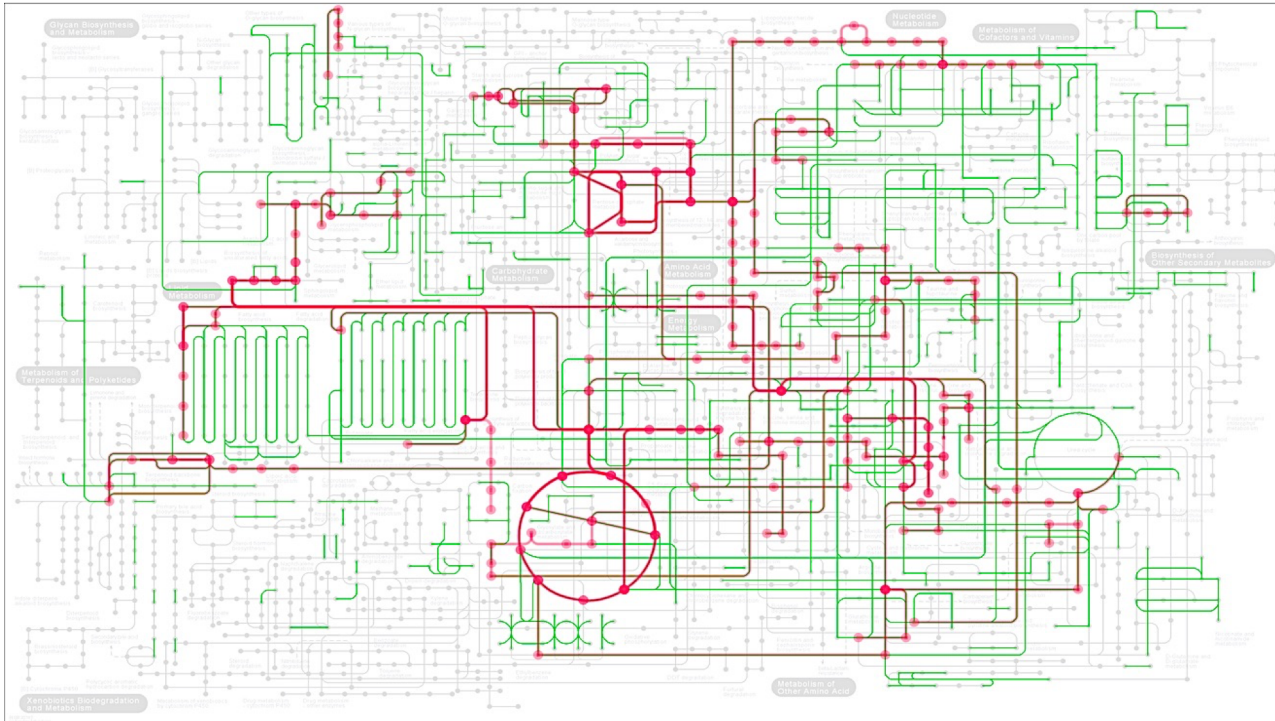


Figure 2.10 Two views of KEGG maps of metabolic pathways present in *Peltigera membranacea*.
 (a) black lines, all pathways represented (b) green lines, all pathways represented and in red, pathways found to be complete.

2.3.5 Identification of Differentially Expressed Genes (DEGs)

After preliminary quality control (QC) of the RNA-Seq reads, the reads from the three tissues; rhizines, apothecia, and thallus were mapped to the annotated *P. membranacea* genome (P.mem v1.1, reference genome). The data was then quantified, and a cumulative distribution plot was drawn to visualize and compare the distribution of expression values in the different datasets. Ideally, datasets showing the same distribution of values (even if individual measures show large differences) should be virtually identical on this plot, and normalizing the data makes them as comparable as possible. All tissues analyzed showed some irregularities in the lower end due to very low absolute counts, but low expression is largely excluded from the general analysis. The expression values were adjusted for all the samples by matching count distributions at the 75th percentile, yielding an equal uniform distribution. Cumulative distribution plots for the thallus whole samples are shown in Figure 2.11.

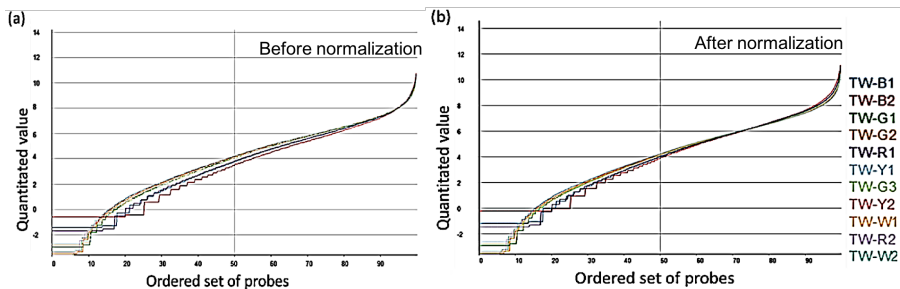


Figure 2.11 Cumulative distribution (CD) plots of thallus whole technical replicates collected in summer 2012.

x-axis shows the ordered set of mRNAs (probes), and y-axis shows the quantitated read counts of the mRNAs (probes) (a) CD plot before normalization (b) CD plot after normalization (at 75th percentile).

Table 2.5 Summary of DEGs.

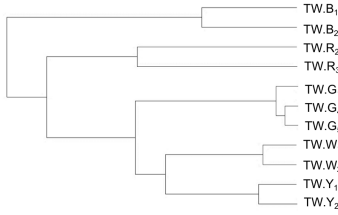
S. NO	Tissue combinations	Before >2X BG variation removal	Number of >2X BG variation removed	After >2X BG variation removal	Appendix
1	Apothecia (autumn 2009) vs rhizine (autumn 2010 and summer 2012) vs thallus body (summer 2012))	413	74	339	M
2	Apothecia (autumn-2009) vs Thallus Whole (autumn-2009)	265	56	209	N
3	Apothecia (autumn-2009) vs Rhizine (autumn-2010)	219	0	219	O
4	Rhizine (autumn-2010) vs Thallus whole (autumn-2010)	183	65	118	P
5	Apothecia (autumn-2009) vs Thallus Body (summer-2012)	84	11	73	Q
6	Rhizine (summer-2012) vs Thallus Body (summer-2012)	294	63	231	R
7	Thallus whole (autumn 2009 vs 2010) vs Thallus Whole (summer 2011 vs 2012)	317	105	212	S
8	Rhizine (autumn-2010) vs Rhizine (summer-2012)	76	0	76	T
9	Thallus Body (summer-2012) vs Thallus Whole-G (summer 2012)	34	7	27	U

720 genes with variable background expression in thallus whole biological replicates i.e. BKG genes are removed from tissue comparisons 1, 2, 4 to 7 and 9

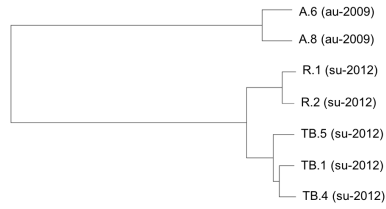
To calculate the correlation value for all pairwise comparisons of the samples, a dendrogram was constructed to show the overall differences between expressed transcripts of technical and biological replicates using the datastore tree option in SeqMonk. The datastore tree clustering is based on a distance matrix comparing every sample with every other sample and uses the neighbor-joining method to construct a tree from the matrix. In the dendrogram, the technical replicates grouped with their respective biological samples (Figure 2.12), showing lower variability than the biological replicates. The technical replicates were pooled for each sample in further analysis. PCA (Principal component analysis) plots were created to visualize the correlation between the quantitated values for all the datasets.

Darastore tress

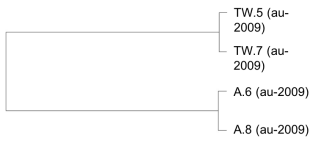
(i) TW-B vs TW-G vs TW-W vs TW-R and TW-Y (su 12)



(ii) A (au-09) vs R (su-12) vs TB (su-12)



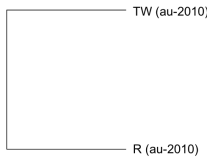
(iii) A (au-09) vs TW (au-09)



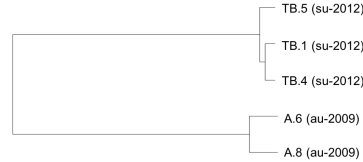
(iv) A (au-09) vs R (au-10)



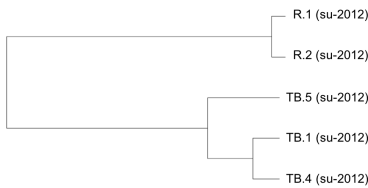
(v) R (au-10) vs TW (au-10)



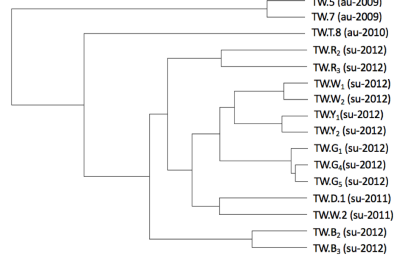
(vi) A (au-09) vs TB (su-12)



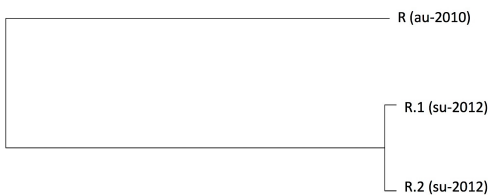
(vii) R (su-12) vs TB (su-12)



(viii) TW (au-09) vs TW (au-10) vs TW (su-11) vs TW (su-12)



(ix) R (au-10) vs R (su-12)



(x) TB (su-12) vs TW-G (su-12)

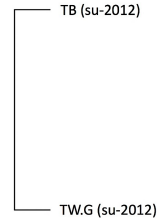


Figure 2.12 Datastore tree of the different datasets. The relationship of the technical replicates and the respective biological replicates are shown.

Differential gene expression for the comparisons presented in Table 2.5 was carried out. The list of DEGs generated using the intensity difference method was saved for further analysis. To aid the interpretation of DEGs results, hierarchical clustering was done and heatmaps were generated (Figure 2.13, Figure S21).

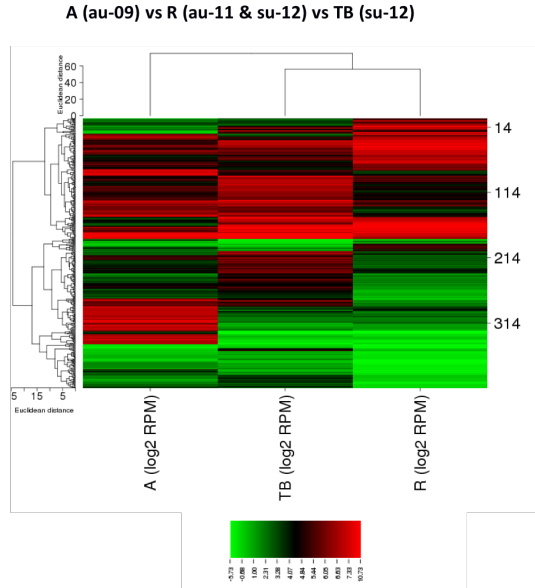


Figure 2.13 Heatmap from hierarchical clustering of Apothecia (autumn 2009) vs. rhizine (autumn 2010 and summer 2012) vs. thallus body (summer 2012).

Rows correspond to genes and columns to samples. Intensities recorded for genes are colour coded: Green-low expression, Red-high expression, Black-medium expression.

The summary of DEGs from the nine tissue comparisons before and after subtraction of background (BKG) genes (See section 2.2.6) is presented in Table 2.5. Most of the DEGs were hypothetical, uncharacterized or predicted proteins (Figure 2.14). Although many of the DEGs did not yield significant BLASTP results, there were some interesting first insights from these analyses. DEGs included the heat shock proteins (HSP) which are stress-related proteins (Paoletti and Clavé 2007; Tiwari et al. 2015). In *P. membranacea*, *hsp30* showed a two-fold higher expression in apothecia compared to thallus and rhizine (Table 2.6). *hsp20* and *hsp70* were differentially expressed in autumn 2009, 2010 samples (Table 2.6) and summer 2012 biological replicates (TW-B vs. G vs. R vs. W vs. Y) (Table S18), respectively. Three

cytochrome P450 genes were highly expressed (>2X fold) in thallus than the mycobiont tissues (Table 2.6) and differentially expressed in the thallus biological replicates (Table 2.6).

Table 2.6 List of *hsp* and cytochrome P450 genes in *P. membranacea*.

Tissue comparisons	Gene id	Description	A (au-09)	TW (au-09)	TW (au-10)	TW (su-11)	TB (su-12)	R (su-12)	TW (su-12)
A (au-09) vs. TB (su-12) vs. R (su-12)	g11615	Heat shock 30	7.32	-	-	-	4.59	4.22	-
	g8665	Heat shock 70	1.14	-	-	-	-0.73	-0.71	-
A (au-09) vs. TW (au-09)	g11615	Heat shock 30	7.32	3.49	-	-	-	-	-
TW (au-09) vs. TW (au-10) vs. TW (su-11) vs. TW (su-12)	g174	Heat shock 30	-	6.81	3.45	6.81	-	-	3.23
	g648	Heat shock 20	-	7.63	4.60	7.45	-	-	7.08
A (au-09) vs. TB (su-12) vs. R (su-12)	g4702	Cytochrome P450	4.36	-	-	-	4.45	1.81	-
	g9085	Cytochrome P450	2.66	-	-	-	3.27	-1.43	-
	g11483	Cytochrome p450	3.29	-	-	-	5.11	0.89	-
	g11486	Cytochrome p450	3.52	-	-	-	5.15	1.60	-

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change. '-' not applicable.

A detailed representation of the various GO terms at hierarchy level 2 for the three categories, Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) is presented in Figure 2.16. GO analyses in the BP category indicated that most of the DEGs were involved in metabolic, single-organism, cellular processes, and localization. DEGs of the MF and CC categories were involved in the catalytic activity, binding and membrane and cell part, respectively. Enzyme commission (EC) numbers associated with the DEGs were also extracted. Most of these DEGs were found associated with enzyme families such as hydrolases, oxidoreductases, ligases, lyases, and transferases (Figure 2.17). No differentially expressed enzyme transcripts were found for TB (su-12) vs. TW-G (su-12).

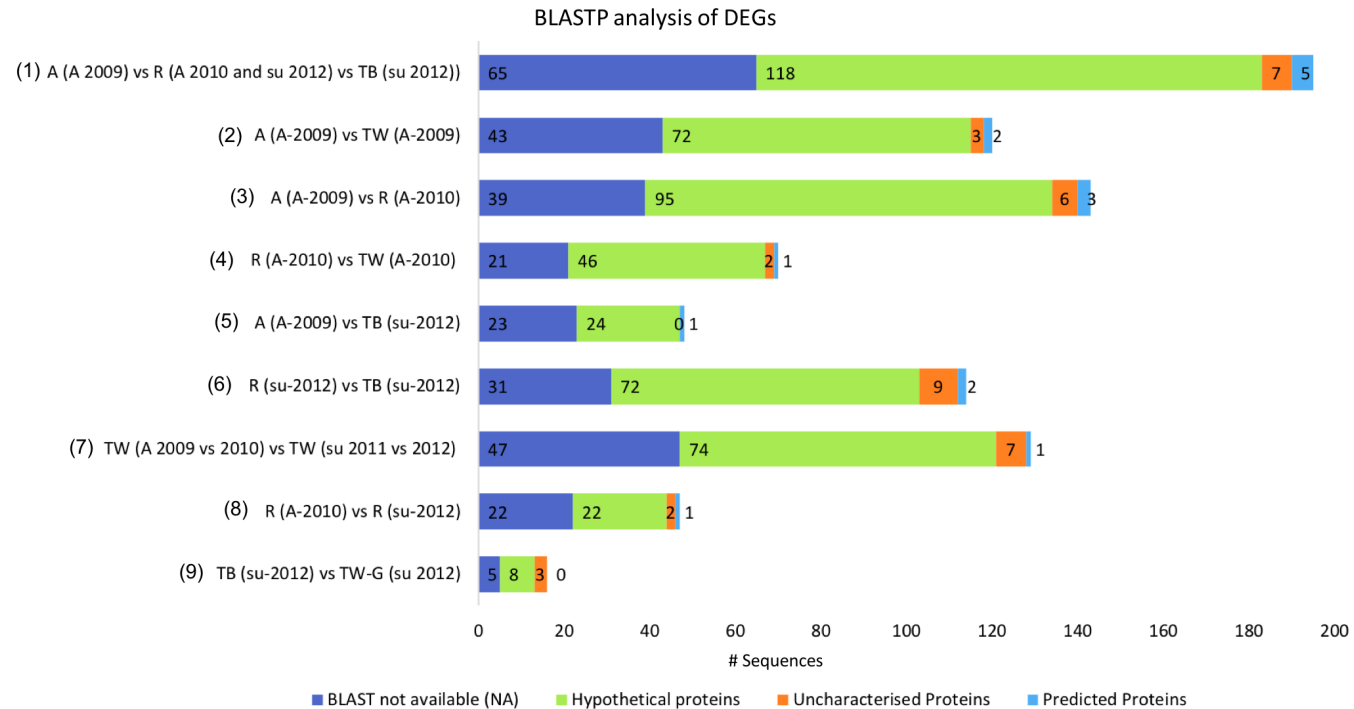


Figure 2.14 BLASTP analysis of DEGs.

Predicted proteins are used for entries without evidence at protein, transcript or homology levels. (BKG genes are removed from tissue comparisons 1, 2, 4 to 7 and 9).

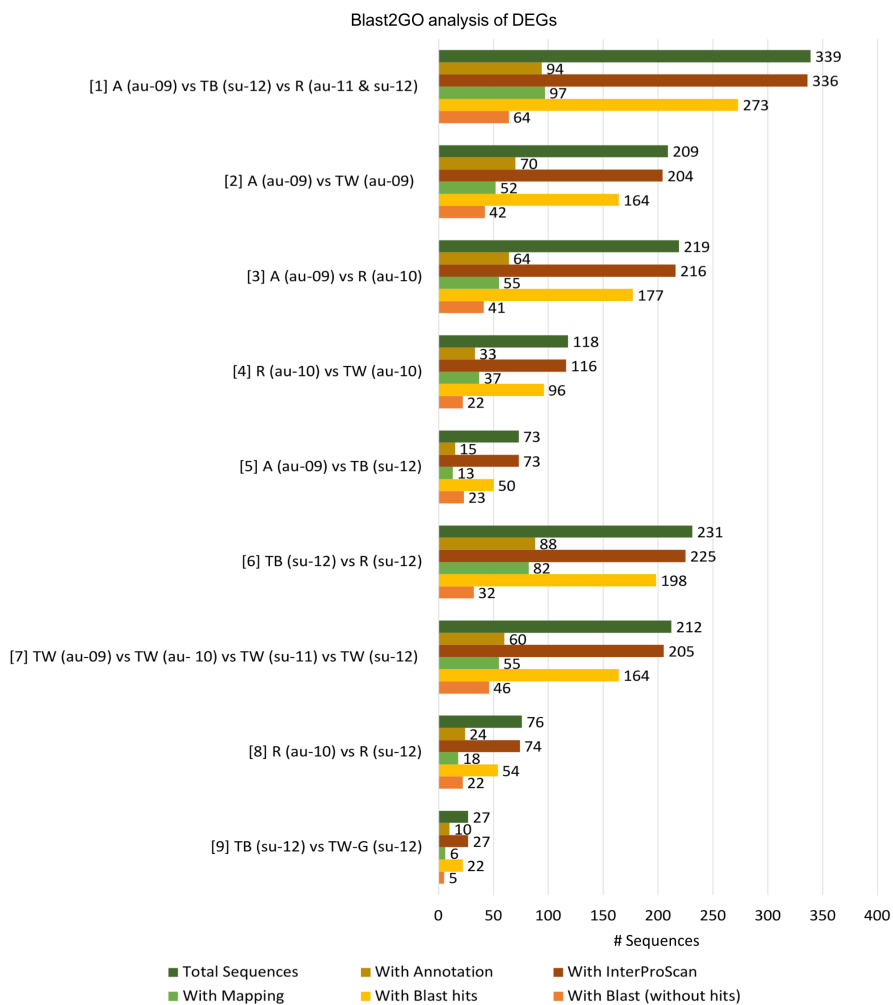
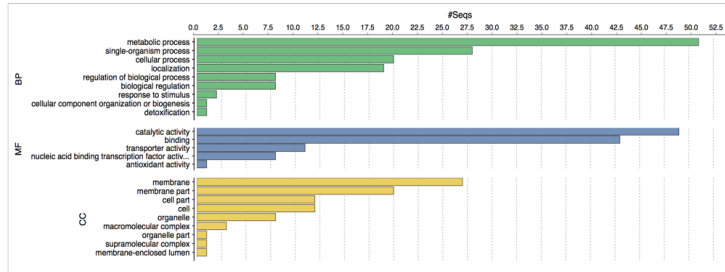
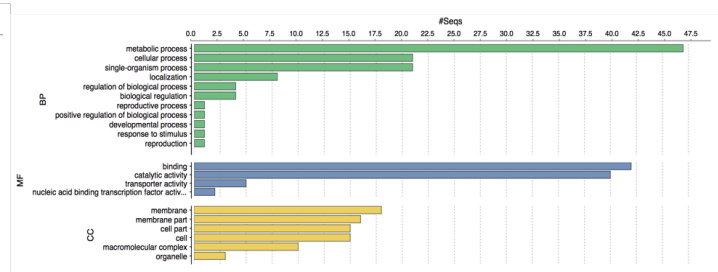


Figure 2.15 Blast2GO analysis of DEGs.
(BKG genes are removed from tissue comparisons 1, 2, 4 to 7 and 9)

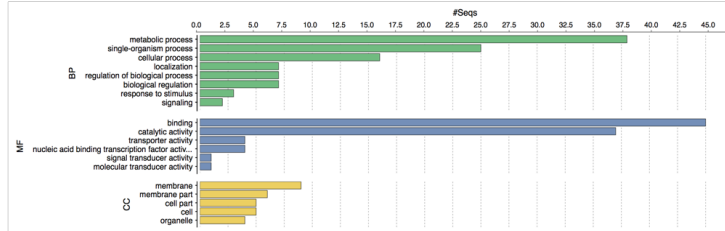
[1] A (au-09) vs TB (su-12) vs R (au-11 & su-12)



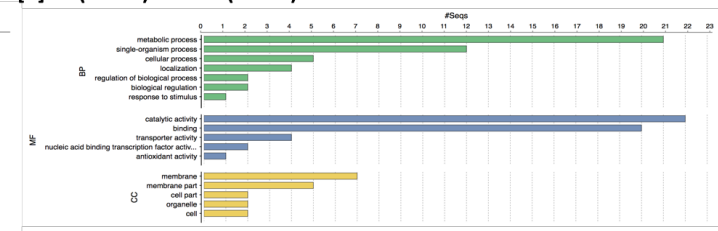
[2] A (au-09) vs TW (au-09)



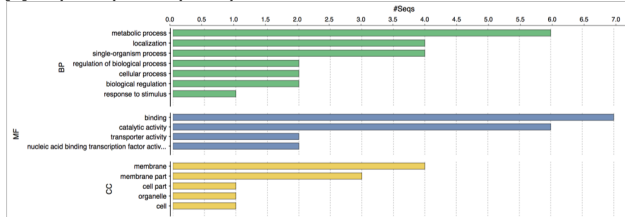
[3] A (au-09) vs R (au-10)



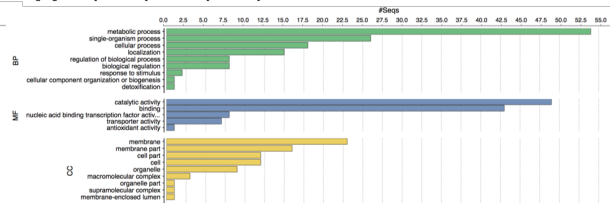
[4] R (au-10) vs TW (au-10)



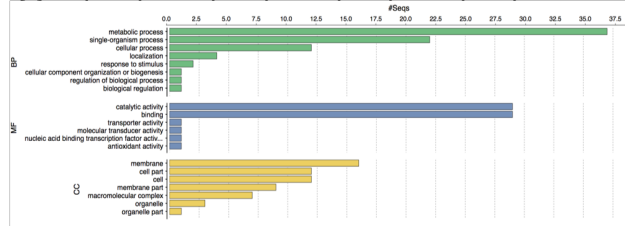
[5] A (au-09) vs TB (su-12)



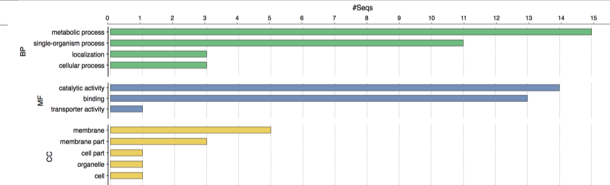
[6] TB (su-12) vs R (su-12)



[7] TW (au-09) vs TW (au-10) vs TW (su-11) vs TW (su-12)



[8] R (au-10) vs R (su-12)



[9] TB (su-12) vs TW-G (su-12)

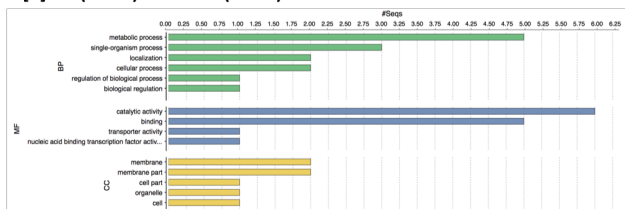


Figure 2.16 Bar Chart of GO annotation of DEGs annotated to the different GO terms at hierarchy level 2 for all the three GO categories.

Biological Process (BP) shaded in green, Molecular Function (MF) shaded in blue and Cellular Component (CC) shaded in yellow in the different tissue comparisons.

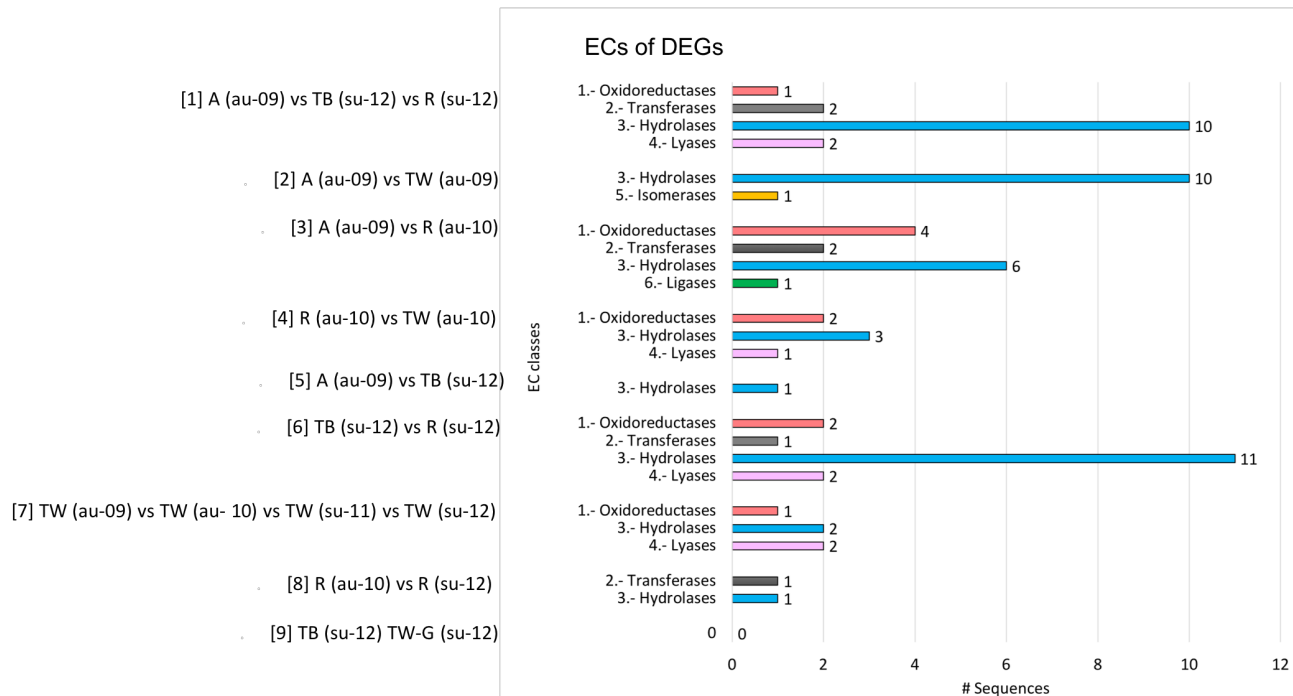


Figure 2.17 Enzyme classification for DEGs.
(BKG genes are removed from tissue comparisons 1, 2, 4 to 7 and 9).

2.3.6 Analysis of specific genes and pathways

To detect tissue specific expression of metabolic activity, the RNA expression data from TB (su-12), R (su-12) and A (au-09) were used. Apothecia were either available at low frequency or not present throughout all seasons, and therefore the only available apothecia data from au-2009 was used. The study focused on key genes involved in nutrient assimilation, the melanin pathway, and genes thought to be involved in symbiosis and genome defense (see Chapter 3). The principal components of these metabolic pathways and their expression are described below.

2.3.6.1 Nutrient assimilation in *P. membranacea*

2.3.6.1.1 Carbon

Four glucose, one sucrose, four monosaccharide, and 12 sucrose transporters were identified. Out of the total 21 transporters, 14 were differentially expressed (Table 2.6). Two enzymes involved in mannitol metabolism in fungi are mannitol-1-phosphate 5-dehydrogenase (MPD; EC 1.1.1.17) and mannitol-1-phosphate phosphatase (MPP; EC 3.1.3.22). Two MPD genes were found to be highly expressed in all tissues but MPP could not be identified. Glucokinase and phospho-glucose isomerase, enzymes involved in the conversion of glucose to glucose 6-phosphate and fructose 6-phosphate, respectively, were also expressed in all tissues (Table 2.6).

2.3.6.1.2 Nitrogen

In cyanolichens N_2 is fixed in the cyanobiont and the reduced nitrogen is transferred to the mycobiont (Rai et al. 1980; Millbank and Kershaw 1969; Englund 1977). Two pathways that are involved in NH_4^+ assimilation are present in *P. membranacea* (Table 2.6, Figure 2.19). The glutamine synthetase (GS) – glutamine oxoglutarate aminotransferase (GOGAT) pathway, also known as the GS-GOGAT pathway, is activated under nitrogen starvation. In this pathway, the high-affinity ammonium assimilating enzyme GS catalyzes the condensation of glutamate (Glu) and ammonia, with the help of ATP, to form glutamine (Gln). Then, Gln and 2-oxoglutarate are converted into 2xGlu by GOGAT. In the NAD(P)-glutamate dehydrogenase (GDH) pathway, Glu is produced via assimilation of NH_4^+ and 2-oxoglutarate by GDH (Botton and Chalot 1999).

GOGAT (EC 1.4.1.14) was upregulated in apothecia and rhizines compared to tissue with cyanobionts (thallus tissue). Glutamine synthetase (GS, EC 6.3.1.2), NADP⁺ dependent glutamate dehydrogenase (GDH, EC 1.4.1.4) and NAD⁺ dependent glutamate dehydrogenase (GDH, EC 1.4.1.2) showed similar expression in all three tissues. Several other transcripts directly or indirectly involved in nitrogen metabolism were also identified. These include glutamate oxaloacetate transaminases (SGOT, EC 2.6.1.1; alternative name: aspartate aminotransferases, AAT), glutamate pyruvate transaminase (GPT; alternative name: alanine aminotransferase AlaAt, EC 2.6.1.2) and carbamoyl phosphate synthetase (CPSI, EC 6.3.4.16). Gln acts as a precursor in the synthesis of other amino acids, primarily via transamination.

Amino acid transporters and amino acid permeases identified in *P. membranacea* are listed in Table 2.6 along with the expression level in each tissue. Also, three enzymes which mediate the last step of amino acid biosynthesis pathways, that is: cysteine synthase (EC 2.5.1.47), threonine synthase (EC 4.2.3.1) and methionine synthase (EC; 1.16.1.8), the last one showing high expression in all three tissues.

Table 2.6 List of genes and enzymes encoded in *P. membranacea* involved in nitrogen and carbon metabolism.

EC No.	Enzyme	Symbol	Gene id.	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
Carbon assimilation						
5.3.1.9	Glucose-6-phosphate isomerase (Phosphoglucose isomerase)	GPI (PGI)	g4152	7.97	7.64	7.76
2.7.1.2	Glucokinase	GLK1	g9511	6.63	6.12	5.45
	Glucose transporter		g10695	7.18	6.74	6.96
	MFS glucose transporter		g2301	2.80	2.45	1.62
	MFS glucose transporter		g7386	2.89	4.64	4.69
	High-affinity glucose transporter		g9131	3.21	4.38	3.63
	MFS monosaccharide transporter		g9455	6.03	5.73	5.80
	MFS monosaccharide transporter, partial		g5770	5.44	5.28	4.54
	Monosaccharide transporter		g4883	3.22	-2.41	-2.77
	Monosaccharide transporter		g5619	8.07	7.42	7.23
	MFS sugar transporter		g1183	4.93	1.84	1.64
	MFS sugar transporter		g15114	0.47	2.92	1.23
	MFS sugar transporter		g1669	7.18	7.59	7.69
	Sugar transporter		g2017	3.24	3.48	3.41
	Sugar transporter		g2056	5.09	3.18	3.36
	Sugar transporter		g6025	1.51	-0.79	-0.76
	Sugar transporter STL1		g6278	4.83	6.30	4.70
	Sugar transporter conserved site		g9525	6.02	5.48	5.52
	Sugar transporter		g10495	-0.10	-1.40	-1.37
	Sugar transporter		g11167	5.15	5.10	5.02
	Sugar transporter		g12661	4.34	5.90	6.00
	Sugar transporter		g1510	8.08	7.94	8.26
	Sucrose transporter		g3161	5.57	5.29	5.38
1.1.1.17	Mannitol-1-phosphate dehydrogenase	MPD	g212	5.90	5.29	5.01
			g5696	8.39	8.57	8.49
Nitrogen assimilation						
2.6.1.1	Glutamate oxaloacetate transaminase (aspartate aminotransferases)	SGOT (AAT)	g171	3.52	3.15	4.11
			g7577	5.75	5.42	5.77
	Putative aminotransferases	AAT-like	g7576	6.17	6.71	6.93
	Similar to aspartate aminotransferase		g14587	5.99	6.15	6.59
6.3.1.2	Glutamine synthetase	GS	g4938	7.83	7.28	7.68
			g13236	6.10	6.64	6.66
1.4.1.14	Glutamate synthase (NADPH)	GOGAT	g1326	6.92	5.83	6.71
	Glutamate synthase (NADH)		g1327	7.52	6.24	7.04
1.4.1.4	Glutamate dehydrogenase (NADPH)	gdhA	g10387	6.80	6.75	6.79
			g10388	6.51	6.64	6.60
1.4.1.2	Glutamate dehydrogenase (NADH)	GDH	g7052	5.69	5.25	5.27
2.6.1.2	Glutamate pyruvate transaminase (alanine aminotransferase)	AlaAT (GPT)	g1508	6.81	6.65	6.99
6.3.4.16	Carbamoyl phosphate synthetase	CPSI	g9502	6.47	5.58	5.44
2.6.1.44	Alanine-glyoxylate aminotransferase	AlaGAT	g526	5.61	5.92	5.98
2.5.1.47	Cysteine synthase	CysK	g2428	6.32	5.55	5.56
1.16.1.8	Methionine synthase	MS	g5737	7.69	7.81	7.53
4.2.3.1	Threonine synthase	TS	g13721	5.32	5.02	4.98
			g7690	6.67	6.16	6.73
1.A.11.3.2	High-affinity ammonium transporter	MEP2	g16330	6.65	5.78	6.72
			g14188	6.69	5.75	6.87
	Low-affinity ammonium transporter	MEP3	g8261	5.22	6.73	6.17
	Nitrate transporter		g4757	4.90	5.73	5.98
1.7.1.4	Nitrite reductase		g6402	2.94	5.28	5.27
1.7.99.4	Nitrate reductase (NADPH)		g6403	3.78	6.09	6.15

Amino acid transporter		g3008	5.64	5.13	5.17
Amino acid transporter		g9719	7.51	8.02	8.10
Amino acid transporter		g13942	4.97	6.11	6.17
Amino acid transporter		g16063	4.53	4.69	4.17
Amino acid transporter		g12771	5.80	6.75	4.81
Amino acid transporter arg-13		g16127	4.13	5.60	3.91
aat family amino acid transporter		g12509	6.96	7.27	7.67
Similar to amino acid transporter		g81	4.17	4.53	4.49
Similar to amino acid transporter		g3168	6.60	6.72	5.27
Similar to amino acid transporter		g9229	2.99	0.82	1.58
Vacuolar amino acid transporter 5		g13129	3.73	4.29	3.99
Putative vacuolar amino acid transporter YPQ3		g9315	5.58	2.69	2.05
Amino acid transporter, transmembrane		g2295	0.84	0.64	-0.69
Amino acid permease (Gap1)	GAP1	g4311	5.87	5.84	5.90
apc amino acid permease	APC	g7741	4.68	5.81	5.97
gaba permease	GABA	g13786	4.16	3.97	5.96
Vacuolar amino acid permease		g10051	3.57	1.69	2.54
Nitrogen permease regulator 2		g8824	5.18	4.46	4.37

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change.

2.3.7 Transporters in *P. membranacea*

Transporters are involved in controlling the flux of molecules for generation of energy, delivery of nutrients and elimination of metabolic waste from the cell (Raethong et al. 2016). Transporter proteins for ammonium, magnesium, copper, carbohydrate, and amino acid transporters were identified in the annotated genome of *P. membranacea* (Table 2.7). NH_4^+ is taken up by the ammonium transporter/methyl ammonium permease family (MEP) of transporter proteins (Ninnemann et al. 1994). Three high-affinity ammonium transporter (*mep2*) transcripts were expressed in all three tissues, with the lowest expression level in the thallus body. A single low-affinity ammonium transporter (*mep3*) transcript was found to be most strongly expressed in the thallus body followed by rhizine and apothecial tissues (Table 2.7).

Two major classes of transporter proteins, the ATP-binding cassette (ABC) transporters and the major facilitator superfamily (MFS) multidrug transporters that play a major role in fungicide sensitivity and resistance (Del Sorbo et al. 2000) were identified. Two ABC multidrug transporters and 12 MFS multidrug transporters are present in *P. membranacea* (Table S40). Four zinc transporters were expressed in all the tissues suggesting zinc tolerance and accumulation in *P. membranacea*, as was seen in *A. oryzae* (Raethong et al. 2016). The transporters identified in *P. membranacea* are listed in Table S40.

2.3.8 Melanin pathway

Melanin provides defense against environmental stress such as UV radiation, oxidizing agents and ionizing radiation (Nosanchuk and Casadevall 2003; Jacobson 2000). In lichens, melanin is widely distributed (Chen et al. 2014; Beckett et al. 2012). The mycobiont synthesizes melanins to protect itself and its photobiont partner (Galun 1998). In fungi, three different pathways can contribute to melanin production; the DHN–melanin pathway; the DOPA–melanin pathway; and the L–tyrosine degradation pathway (Chen et al. 2014). In *Aspergillus*, the DHN–melanin pathway includes a polyketide synthase (PKS), scytalone dehydratase (ARP1), 1,3,6,8–tetrahydroxynaphthalene reductase (ARP2), an alpha/beta hydrolase (AYG1), and two multicopper oxidases (MCOs), a ferroxidase (ABR1), and a pigment MCO (ABR2) (Table 2.7). Gene homologs for all these proteins were identified in *P. membranacea*, and the PKS (g4511_scaff38861) and ABR1 (ferroxidase) (g10570_scaff30968) were highly upregulated in rhizines and in general the DHN melanin pathway genes were upregulated in rhizines and downregulated in apothecia. The DOPA–melanin pathway genes were also present, with both laccases and tyrosinases. Tyrosinase genes were expressed in apothecia and out of the four tyrosinases, one was upregulated (about two–fold). Three laccase genes were identified, and two of them showed significant expression in all three tissues. An alternative pathway to produce a third type of melanin, using the L–tyrosine degradation pathway, is also conserved in *P. membranacea* (Table 2.7). The genes concerned were expressed at low levels.

Table 2.7 List of genes involved in the melanin biosynthesis pathway.

Gene	<i>A. fumigatus</i>	<i>A. niger</i>	Description	<i>P. membranacea</i> top hit with lowest e value	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
DHN-melanin pathway							
Pks1	Afu2g17600	An03g05440	Polyketide synthase	g4511_scaff38861 (0)	5.71	6.48	8.63
	Afu4g00210	An04g09530					
	Afu4g14560	An09g05730					
	Afu7g00160	An11g07310					
Ayg1	Afu2g17550	An14g05350 (Ayg1)	Abhydrolase	g10570_scaff30968 (6E-165)	3.28	4.74	6.17
Arp2	Afu2g17560	An02g00220	1,3,6,8-Tetrahydroxynaphthalene	g4949_scaff30704 (7E-55)	4.20	4.86	4.72
Arp1	Afu2g17580	An08g09920	Scytalone dehydratase	g10568_scaff30968 (4E-23)	4.31	5.99	6.56
Abr2	Afu2g17530	An01g13660	Fungal pigment MCO	g2620_scaff30629 (2E-27)	5.06	4.42	5.25
	Afu1g15670	An01g14010		g8781_scaff38132 (5E-43)	0.77	3.33	5.50
		An03g03750					
		An04g10400					
Abr1		An05g02540	Fungal ferroxidase				
	Afu2g17540	An14g05370 (BrnA)		g530_scaff30992 (1E-168)	3.92	6.12	8.50
	Afu5g03790 (FetC)	An01g11120					
		An01g08960					
		An15g05520					
DOPA-melanin pathway							
melC2	Afu3g01070	An01g09220	Tyrosinase	g2146_scaff31840 (2E-75)	4.22	3.25	1.81
melO		An03g00280		g5550_scaff33648 (3E-85)	2.68	2.83	2.52
		An12g01670		g1251_scaff30955 (2E-18)	8.40	7.10	7.39
		An09g02980		g9463_scaff31764 (2E-72)	1.66	1.23	1.03
Afu4g14490		An12g05810 (McoJ)	Laccase	g3439_scaff30511 (2E-149)	7.32	8.22	7.57
		An16g02020		g7226_scaff32618 (3E-144)	7.12	6.54	6.76
		An11g03580					
		An08g08450					
		An05g02340		g2620_scaff30629 (1E-116)	5.06	4.42	5.25
		An01g00860					
		An18g02690 (McoJ)					
L-tyrosine degradation pathway							
Tat	Afu2g13630	An02g05540	Tyrosine aminotransferase	g9498_scaff38795 (0)	5.03	4.52	4.15
hpdD	Afu2g04200	An11g02200	4-Hydroxyphenylpyruvate dioxygenase	g5509_scaff36848 (7E-160)	4.94	4.52	4.56
hmgA	Afu2g04220	An11g02180	Homogentisate dioxygenase	g10765_scaff31816 (0.9)	5.17	0.76	0.27
fahA	Afu2g04230	An11g02170	Fumarylacetoacetate hydrolase	g10476_scaff34199 (6.3)	4.57	4.11	3.99
maiA	Afu2g04240	An11g02160	Maleylacetoacetate isomerase	g8195_scaff32916 (4.00E-04)	5.89	5.37	5.48

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change.

2.3.9 Glyoxylate pathway

The glyoxylate cycle, a ramification of the tricarboxylic (TCA) cycle, was identified in *P. membranacea* (Figure 2.18). The pathway is characterized by two key enzymes: isocitrate lyase (ICL) and malate synthase (MLS) (highlighted in red in Figure 2.18). The former was found to be upregulated in apothecia (Table 2.8). TCA cycle (Figure 2.18) genes along with expression values are listed in Table 2.8.

Table 2.8 List of genes and enzymes encoded in *P. membranacea* involved in glyoxylate cycle.

Gene	EC No.	Gene id.	Description	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
ICL	EC:4.1.3.1	g9221	Isocitrate lyase	5.44	4.67	4.41
		g964	Isocitrate lyase	5.53	5.52	5.48
		g4941	Isocitrate lyase	6.96	6.91	6.92
MLS	EC:2.3.3.9	g2413	Malate synthase	6.41	5.98	5.90
		g14889	Malate synthase A	4.58	3.71	3.68
CS	EC:2.3.3.1	g963	Citrate synthase	5.17	4.92	4.91
		g3632	Citrate synthase (<i>Cit1</i>)	6.81	6.86	6.87
ACO1	EC:4.2.1.3	g4707	aconitate hydratase (aconitase)	1.28	3.44	2.64
		g13618	aconitate hydratase (aconitase)	5.82	4.97	4.85
		g5391	aconitate hydratase 1 (aconitase)	6.46	7.63	7.71
		g10800	aconitate hydratase 1 (aconitase)	5.61	6.96	7.19
		g15154	aconitate hydratase 1 (aconitase)	5.53	6.45	6.47
IDH1	EC:1.1.1.42	g3745	Isocitrate dehydrogenase 1 (NADP ⁺)	5.90	5.80	5.83
IDH3	EC:1.1.1.41	g12337	Isocitrate dehydrogenase (NAD ⁺)	7.04	6.57	6.60
		g1720	Isocitrate dehydrogenase (NAD ⁺)	6.73	6.40	6.58
	EC 1.2.4.2	g7033	oxoglutarate dehydrogenase (succinyl-transferring), E1 component	6.79	6.30	6.33
SUCLG1	EC 6.2.1.5	g2970	succinyl-CoA ligase beta-chain (succinyl-CoA synthetase)	6.83	6.50	6.55
		g5233	alpha subunit of succinyl-CoA ligase (succinyl-CoA synthetase)	6.62	6.65	6.81
SDHB		g2826	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	6.92	6.91	7.09
SDHC	EC 1.3.5.1	g5272	succinate dehydrogenase (ubiquinone) cytochrome b560 subunit	6.71	6.54	6.77
SDHD		g55	succinate dehydrogenase-ubiquinone-cytochrome b small subunit, mitochondrial precursor	6.67	6.30	6.51
FH	EC:4.2.1.2	g10118	fumarate hydratase 1, mitochondrial (Fumarase)	6.06	5.44	5.28
MDH	EC:1.1.1.37	g2107	malate dehydrogenase	6.55	6.64	6.77
		g6522	malate dehydrogenase	7.05	7.24	7.32

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change.

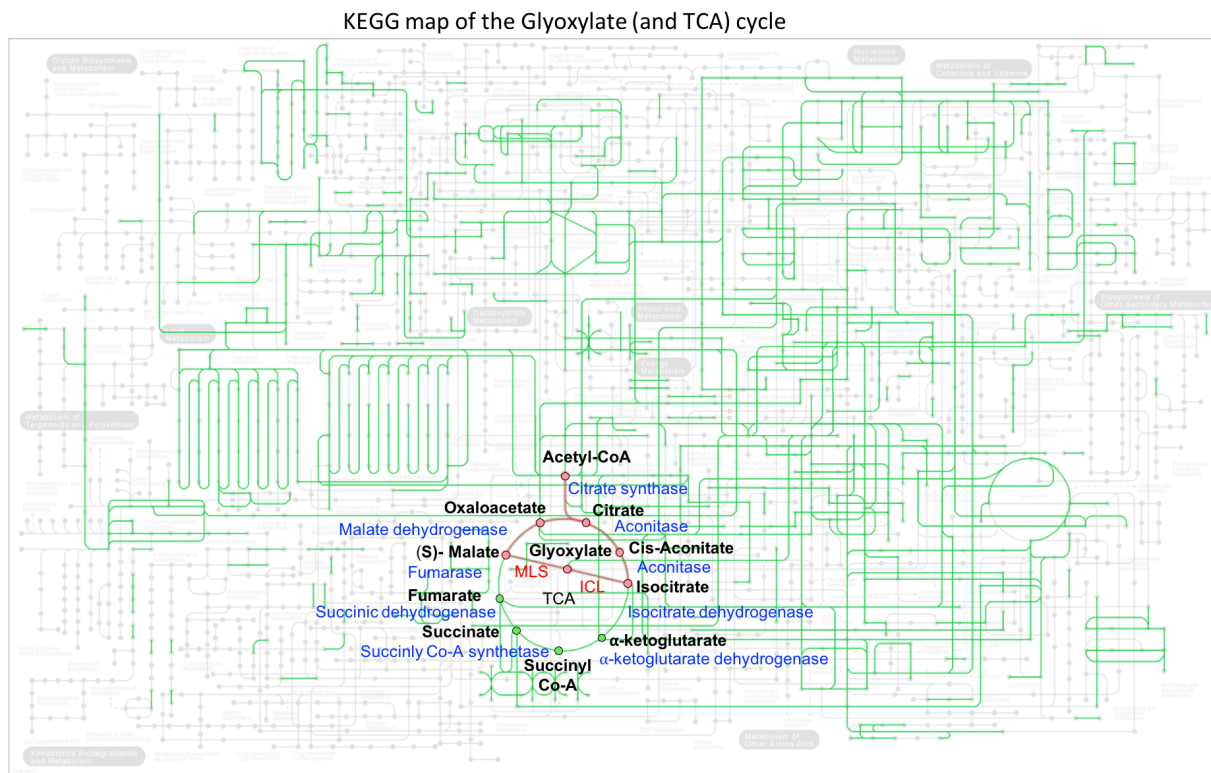


Figure 2.18 KEGG map of the Glyoxylate (and TCA) cycle identified in *Peltigera membranacea*. All pathways that are represented in *P. membranacea* are highlighted in green. The glyoxylate pathway and the two key enzymes; isocitrate lyase (ICL) and malate synthase (MLS) are highlighted in red.

2.3.10 Genes implicated in symbiosis

A search was made for homologs of genes implicated in symbiosis. Twelve genes encoding lectins, four encoding mitogen-activated kinases (MAPK), a hydrophobin gene, and a hydantoin racemase were identified from BLASTP searches of the translated CDS (Table 2.9). Two galectin-like genes, *lec-1* (Miao et al. 2012) and *lec-2* (Manoharan et al. 2012) had been identified previously. *lec-1* and *lec-2* were highly expressed in the thallus body and apothecia, respectively. Additionally, five fucose-lectin genes, two concanavalin A-like lectins, two lectin family integral membrane proteins, and a legume-like lectin were identified. The expression data is presented in Table 2.9.

The concanavalin A-like lectin glucanase gene and a putative fucose-specific lectin gene were found to be highly expressed in thallus and rhizine tissues whereas another fucose-lectin gene was found to be highly expressed in mycobiont tissues (Table 2.9). The hydrophobin H1 gene, also implicated in symbiosis (Scherrer et al. 2002; Trembley et al. 2002) was found to be strongly expressed in the thallus body.

Table 2.9 List of symbiosis-related genes in *P. membranacea*.

Gene id	Description	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
g9266	Concanavalin A-like lectin	1.23	1.70	1.02
g13191	Concanavalin A-like lectin glucanase	4.46	5.98	5.95
g9268	Fucose-specific lectin	4.05	3.38	3.82
g4473	Fucose-specific lectin	1.06	0.85	0.45
g4233	Similar to Fucose-specific lectin (<i>lec-3</i>)	7.30	6.77	7.32
g12994	Similar to Fucose-specific lectin	2.84	8.13	9.40
g16200	Similar to Fucose-specific lectin	7.67	8.63	9.82
g11004	galectin-like protein (<i>lec-1</i>)	3.31	8.43	6.36
g15502	galectin-like protein (<i>lec-2</i>)	7.95	6.12	1.75
g3281	Legume-like lectin family	7.26	6.08	5.60
g2711	Lectin family integral membrane protein	5.71	5.84	6.20
g4800	Lectin family integral membrane protein	6.46	5.96	5.95
g10814	Hydrophobin H1	8.49	10.12	8.79
g4890	Mitogen-activated kinase	6.95	6.76	6.83
g10635	Mitogen-activated kinase	6.91	6.54	6.13
g14647	Mitogen-activated kinase spm1	2.72	1.88	1.51
g3988	Mitogen-activated kinase spm1	7.45	7.12	7.12
g8821	Hydantoin racemase	1.42	2.65	1.42

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change.

2.3.11 Identification of Mating-type and heterokaryon incompatibility genes

Putative *mat* genes encoding alpha and HMG domains were identified by using protein sequences from *A. nidulans* and *N. crassa* (Table 2.10) as queries in TBLASTN and BLASTP (Altschul et al. 1997) searches against the *P. membranacea* genome (P.mem v1.1) and P.mem v1.1 proteins (translated transcripts) databases, respectively. The search resulted in the identification of proteins and scaffolds containing MAT alpha-box (MAT1-1-1) and HMG-box (MAT1-2-1) hits. The identified alpha domain and HMG domain proteins were then used as query sequences to perform BLASTP searches in GenBank, and the top hits were the MAT1-1-1 (CAI59771.1) and MAT1-2-1 (CAI59768.2) of *Xanthoria polycarpa*.

Mat genes are linked to *apn-2*, and *sla-2* genes in dimorphic fungi but the genes were present in different scaffolds in P.mem v1.1 (Table 2.10). Both *mat1-1-1* and *mat1-2-1* have >2-fold higher expression in apothecia than thallus and rhizines, whereas *apn-2* and *sla-2* genes showed the same level of expression in all three tissues (Table 2.10).

Table 2.10 List of mating, heterokaryon incompatibility, vegetative compatibility, and sexuality-related genes of *P. membranacea* genes and their homologs in *Aspergillus nidulans*, *Neurospora crassa* and *Podospora anserina*.

Gene Description	<i>A. nidulans</i>	<i>N. crassa</i>	<i>P. anserina</i>	<i>P. membranacea</i>	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
ABC transporter required for a-factor transport	AN2300	NCU07546	—	g6898_scaff30648_complement (0)	6.52	5.91	6.10
Amino acid permease involved in sexual differentiation	AN5678	NCU05830	—	g4311_scaff33611 (0)	5.87	5.84	5.90
APC component	AN0905	NCU00494	—	g10535_scaff40166 (1.2E-41)	4.32	3.41	3.14
APC component	AN8002	NCU01377	—	g10398_scaff31025_complement (2.7E-169)	5.68	5.17	5.03
APC component	AN2772	NCU05901	—	g13435_scaff45281_complement (0)	5.55	4.56	4.51
APC component	AN4735	NCU01963	—	g5382_scaff31241 (1.16238E-126)	4.90	4.74	4.48
APC component	AN8013	NCU01174	—	g1490_scaff33660 (0)	6.38	6.10	5.96
APC regulator	AN2965	NCU01269	—	g1350_scaff30836_complement (0)	5.50	4.38	4.47
APC regulator	AN0814	NCU02616	—	g6425_scaff30963_complement (0)	7.78	6.48	6.53
DNA lyase, APN-2	—	NCU01961	—	g10886_scaff35023 (2.19E-24)	5.07	5.21	5.16
Ascospore lethal-1	AN2911	NCU01345	—	g13079_scaff31176_complement (9.0E-43)	6.10	7.34	7.40
Ascospore maturation-1	AN5836	NCU01414	—	g13916_scaff36012_complement (2.0E-112)	7.47	7.14	6.55
Ascus development-4	AN6221	NCU07039	—	g15123_scaff40465_complement (6.7E-12)	5.03	6.38	6.30
Ascus development; rhamnogalacturonase B	AN7135	NCU05598	—	g9044_scaff30790_complement (8.31)	5.39	4.92	4.89
ATP-dependent efflux pump for a-factor like pheromone	AN2300	NCU07546	—	g6898_scaff30648_complement (0)	6.52	5.91	6.10
B-type cyclin	AN2137	NCU01242	—	g7479_scaff33964 (1.0E-157)	5.64	4.14	3.54
Binuclear zinc transcription factor	AN5170	—	—	g12392_scaff32281 (2E-165)	6.70	5.70	5.27
CAAX prenyl protease a-factor C-terminal processing	AN6528	NCU11314	—	g10050_scaff33416_complement (1.0E-78)	5.39	5.46	5.73
Chromosome segregation, kinetochore-associated Ndc80 complex	AN4969	NCU03899	—	g2457_scaff41417 (0)	4.36	4.35	4.13
Dipeptidyl aminopeptidase alpha-factor processing STE13	AN2946	NCU02515	—	g191_scaff30559_complement (0)	6.41	6.40	6.59
Dipeptidyl aminopeptidase for a-factor processing STE23	AN8044	NCU00481	—	g1479_scaff33660 (0)	7.07	5.55	5.52
DNA-binding helix-hairpin-helix protein, DNA strand exchange	AN9092	—	—	g821_scaff30748 (0)	5.57	4.98	3.85
Endoprotease for alpha-factor processing	AN3583	NCU03219	—	g13839_scaff33350 (2.4E-151)	6.87	6.56	6.49
Fork head domain TF, meiotic regulator	AN8858	NCU06173	—	g7865_scaff31267_complement (4.3E-40)	5.47	3.94	3.85
GATA-transcription factor	AN3152	NCU01154	—	g5680_scaff33557 (3.7E-07)	6.69	6.25	6.12
GEF involved in conjugation; related to Sc CDC25	AN2130	NCU06500	—	g5320_scaff32387 (0)	6.05	5.52	5.66
GTP binding (alpha-1 subunit) involved in conjugation	AN3090	NCU06729	—	g12565_scaff33773_complement (4.2E-93)	6.43	6.41	6.19
Heterokaryon incompatibility gene, HET-6 , allelic het-gene, TOL/HET-6/HET-E domain	—	NCU03533.2	—	g11695_scaff31555_complement (3E-40)	2.85	1.60	4.95
Heterokaryon incompatibility gene, HET-C , allelic het- gene, signal peptide	—	NCU03125.2	—	g7369_scaff34248_complement (0.0)	7.35	6.76	6.85
Heterokaryon incompatibility gene, HET-C2, non allelic vegetative incompatibility	—	—	AAL37301	g9138_scaff31918 (7E-69)	7.29	6.91	6.90
Heterokaryon incompatibility gene, HET-D2Y, beta transducin-like protein	—	—	AAL37301	g14650_scaff48456 (1E-97)	3.05	3.15	3.04
Heterokaryon incompatibility gene, UN-24, ribonucleotide reductase large subunit	—	NCU03539.2	—	g5163_scaff32396_complement (0.0)	5.89	4.10	4.08
HMG-box TF, target of pheromone signaling	AN3667	NCU09387	—	g8907_scaff31206 (6.5E-21)	3.51	3.52	4.40
Incompatibility related gene, HAM-2 protein required for hyphal anastomosis	—	NCU03727.2	—	g3005_scaff30869 (0.0)	6.13	5.82	5.72
Incompatibility related gene, IDI-6, subtilisin-like serine protease	—	—	AAC03564	g1447_scaff33017_complement (0.0)	8.23	7.81	8.05
Incompatibility related gene, IDI-7	—	—	AAN41258	g9492_scaff51872_complement (7E-81)	8.17	7.77	8.01

Incompatibility related gene, PIN-C allelic gene with HET-domain, linked to <i>het-C</i>	—	NCU03494.2	—	g13544_scaff33628_complement (1E-54)	1.64	1.47	2.27
Inducer of meiosis, S/T kinase	AN6243	NCU01498	—	g1154_scaff30861 (0)	5.51	6.07	5.81
Kinesin-like motor required for karyogamy	AN6340	NCU04581	—	g468_scaff31570 (5.2E-128)	5.50	4.04	3.96
M-factor farnesyl cysteine carboxyl methyltransferase	AN6162	NCU00034	—	g13666_scaff41889 (1.3E-36)	5.04	4.97	5.08
MADS-box domain TF, pheromone receptor activator	AN8676	NCU07430	—	g14024_scaff44901 (1.3E-25)	7.00	6.65	6.74
Mating response protein POI2	—	NCU05768.4	—	g12879_scaff30997 (1.38)	4.58	1.68	2.16
Mating type, MAT1-2-1, HMG	AN4734	NCU01960	—	g14431_scaff41924_complement (1E-21)	6.11	2.22	2.41
Mating-type, MAT1-1-1, alpha	AN2755	NCU01958	—	g13506_scaff31318_complement (9E-18)	3.72	0.71	0.39
Mating-type switching protein swi10	AN4331	NCU07066	—	g15111_scaff52844 (2E-97)	4.05	3.53	3.20
Meiosis-specific transcriptional activator	AN6015	NCU09915	—	g10793_scaff38253_complement (4.9E-59)	8.07	5.77	5.52
Meiotic B-type cyclin	AN2137	NCU01242	—	g7479_scaff33964 (6E-155)	5.64	4.14	3.54
Microtubule-binding protein	AN2862	NCU00243	—	g1011_scaff33439 (5.8E-89)	6.49	6.61	6.33
Modifier gene, MOD-A, modifier of <i>het-c/e</i> , <i>c/d</i> and <i>r/v</i> incompatibility, SH3-binding motif	—	—	AAC25496	g1999_scaff31924 (8E-25)	4.86	4.59	4.56
Modifier gene, MOD-D, modifier of <i>het-c/e</i> , incompatibility, G protein alpha subunit	—	—	AAC24766	g7437_scaff35963_complement (0.0)	6.66	6.85	6.82
Modifier gene, MOD-E, modifier of <i>het-r/</i> , incompatibility, HSP90	—	—	AAB97626	g8045_scaff36124_complement (0.0)	8.26	7.56	7.54
Modulator of pheromone-inducible gene expression	AN8676	NCU07430	—	g14024_scaff44901 (3E-26)	7.00	6.65	6.74
Mutanase, cell wall turnover during sexual development	AN7349	—	—	g1497_scaff33660_complement (2E-63)	6.12	6.02	5.01
Pheromone adaptation feedback response	AN7252	NCU00455	—	g12056_scaff32948 (1.5E-117)	5.76	5.86	6.01
Pheromone precursor (alpha-factor like)	AN5791	—	—	g6030_scaff31847_complement (1E-03)	3.26	-1.69	-1.66
Pheromone Receptor (for a-factor like pheromone)	AN7743	NCU00138	—	g7187_scaff44469_complement (5.3E-09)	3.93	1.47	0.49
Pheromone Receptor (for alpha-factor like pheromone)	AN2520	NCU05758	—	g4109_scaff33264_complement (6.07E-10)	5.99	5.37	5.17
Repressor of b mating type regulated genes	AN8211	NCU01238	—	g10841_scaff38072_complement (0)	5.83	5.37	5.29
Required for synaptonemal complex formation	AN8259	NCU01120	—	g11817_scaff36013_complement (0.0097)	7.85	8.02	8.28
RNA binding protein required for meiotic recombination	AN9090	NCU00768	—	g722_scaff33607_complement (1.6E-146)	7.21	6.59	6.45
RNA-binding protein involved in meiosis	AN6494	NCU00118	—	g12898_scaff39897 (1.6E-22)	6.40	5.62	5.56
RNA-binding protein involved in meiosis, Sp Mei4 target	AN7700	NCU00556	—	g12516_scaff34840_complement (1.3E-83)	5.73	5.28	5.41
SAM domain, similar to Sc STE50	AN7252	NCU00455	—	g12056_scaff32948 (9E-119)	5.76	5.86	6.01
Serine carboxypeptidase, degrades extracellular P-factor	AN2555	NCU06720	—	g13046_scaff47852_complement (0)	6.32	7.20	5.47
Serine/threonine protein kinase, negative regulator of meiosis	AN4935	NCU04990	—	g11856_scaff31679 (1.1E-116)	7.59	7.48	7.85
Signalosome subunit 4, regulation of sexual development CsnD	AN1539	NCU07361	—	g5050_scaff41381 (3.5E-117)	6.02	6.01	5.84
Signalosome subunit 5, regulation of sexual development CsnE	AN2129	NCU00467	—	g1912_scaff33749_complement (3E-133)	6.29	5.66	5.59
SLA-2, cytoskeleton assembly control protein	—	NCU01956	—	g9452_scaff39685 (1.18E-119)	6.40	6.37	6.25
Spindle pole body component	AN5618	NCU09871	—	g1127_scaff30861_complement (1.3E-35)	5.45	5.02	4.93
Sporulation regulated septin	AN4667	NCU02464	—	g7795_scaff30537 (0)	7.52	7.15	7.00
Sporulation-specific enzyme required for spore wall maturation	AN2705	—	—	g110_scaff30559 (3E-20)	4.95	6.59	6.17
Suppressor gene, TOL/HET-6/HET-E domain	—	NCU04453.2	—	g1769_scaff34352 (0.0)	4.89	5.03	4.63
Suppressor gene, VIB-1 regulation of conditiation maybe of nrAPase	—	NCU03725.2	—	g15352_scaff33576 (1E-64)	5.67	6.38	5.01
Velvet activator VeA	AN1052	NCU01731	—	g3799_scaff35731 (4.4E-59)	4.46	4.10	4.57

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change. '—' is blank (not searched).

2.4 Discussion

Current sequencing technology enables us to sequence the whole genome and transcriptome of non-model eukaryotic organisms such as lichens (Junttila and Rudd 2012; Wang et al. 2014; Joneson et al. 2011). RNA-Seq was used to study the tissue-specific gene expression in the three tissues (apothecia, rhizine, and thallus) of *P. membranacea*. As no transcriptome information is currently available for any cyanolichen, this study represents a first key step toward understanding the tissue-specific expression in cyanolichens. To address the tissue-specific gene expression, 1) *de novo* transcriptomes were generated, 2) the genome was annotated with the use of the *de novo* transcriptomes, 3) the completeness of the genome was validated, and 4) genes showing tissue-specific expression were identified. Special attention was given to genes involved in nutrient assimilation, melanin and glyoxylate pathways, symbiosis, and sexual reproduction. The present work paves the way for future functional studies in a hypothesis-driven manner. The analysis provides previously undescribed details on the biological functions expressed within each of the tissues in this lichen.

2.4.1 RNA sequencing

The first challenging task was to separate the three tissues of the lichen, i.e., apothecia, thallus body, and rhizine. Tissues such as rhizines are fragile, making it difficult to separate them from the thallus they are attached to. In order to minimize RNA degradation, the tissues were immersed in RNAlater for transit and during separation. Good quality RNA is crucial for any transcriptome analysis, and lichens contain RNAses as well as secondary metabolites that inhibit the recovery of high-quality RNA when extracted with guanidinium salts (Ghawana et al. 2011; Bugos et al. 1995). Several RNA extraction methods were compared in this study (data not shown) to optimize the RNA extraction from *P. membranacea*. Some RNA degradation in the samples was observed which may have occurred during tissue handling in the laboratory (Figure S5). Although the mRNA was partially degraded it was still usable for library preparation and RNA-sequencing, and the same has been observed for other systems, e.g., pea seeds (Chen et al. 2013) and yeast (Fau and Beker 1986). Furthermore, analysis of mRNA 3' bias across the samples indicated low RNA degradation. In total 25 different libraries from 13 samples, including technical and biological replicates,

were sequenced from the three tissues (for details, see materials and methods). The number of *de novo* assembled transcripts varied for each tissue and this may be due to the read length and coverage, as this is known to influence the number and length of transcripts produced during *de novo* transcriptome assemblies (Conesa et al. 2016; Zhao et al. 2011; Francis et al. 2013).

2.4.2 Estimation of the genome assembly completeness using BUSCO

The results of BUSCO and CEGMA were congruent (Appendix C and D). Presence or absence of the BUSCO genes in the four *de novo* transcriptomes could be influenced by read length and number of reads per sample used in the transcriptome assemblies. The presence of duplicate genes could be due to gene duplication, allelic variation (heterozygosity) in the sample, alternative splicing or misassembly. The former two may contribute to duplication in the genome while the latter, together with tissue-specific expression of isoforms, may contribute to duplicated genes in transcriptomes. Most eukaryotic draft genome assemblies available are either incomplete or contain apparent misassemblies (Salzberg and Yorke 2005). Nonetheless, this first draft genome assembly of *P. mem* v1.1, represents a good-quality assembly as it was missing only a few BUSCO genes and was validated using transcriptomic data and previously annotated genes. It will be a significant resource for further elucidation of the genomic architecture of *P. membranacea* and genome-based studies of this organism.

2.4.3 Genome annotation and validation

Genome annotation of a non-model organism, including the prediction of genes, is always a challenging endeavour. *De novo* transcriptomes can help with gene identification in non-model organisms such as lichens. Also, annotation of metagenomic data can be challenging, and can be greatly improved with support of transcriptomic data. *De novo* transcriptomes were used as ESTs during automatic annotation to predict gene models in *P. membranacea*.

Unexpectedly high numbers of genes predicted from draft genomes, i.e. fragmentation of genes, is also a major problem during genome annotation (Denton et al. 2014). This problem can also be more pronounced in *de novo* transcriptome assemblies due to multiple isoforms and lack of full-length transcripts that arise from low coverage during *de novo* assembly process. For instance, when curated PKS gene clusters

were mapped onto the annotated *P. membranacea* scaffolds (Pmem v1.1), all except PKS6 were fragmented, contributing to the unusually high number of gene models (16,400) in *P. membranacea* (Table S41), whereas the 37.5 Mbp *E. pusillum* (Wang et al. 2014), 34.62 Mbp *C. grayi* (<https://genome.jgi.doe.gov/Clagr3/Clagr3.info.html>), 31.90 Mbp *X. parietina* (<https://genome.jgi.doe.gov/Clagr3/Clagr3.info.html>) genomes consists of 9,285, 11,389, 10,818 gene models, respectively. Furthermore, the reference data set used during automatic annotation using AUGUSTUS would not have been adequate and homologues genes such as the PKS might be missing from the reference data sets. Such errors in the genome annotation can be overcome by: (1) combining data from tissues and samples during the generation of *de novo* transcriptomes (2) including more reference data sets during automated annotation (3) by resequencing with long reads such as PacBio or Oxford Nanopore to improve the genome assembly, and (4) targeted sequencing of the *P. membranacea* genome, as has been done with other fungal genomes such as that of *Botrytis cinerea* (Hahn et al. 2014). Although there are inevitable errors in the genome annotation, nevertheless, the use of *de novo* transcriptomes significantly improved the accuracy of gene prediction and helped predict isoforms and non-canonical splice sites.

2.4.4 Gene ontology (GO) and metabolic pathway analysis

Many gene models with no functional information were found to be expressed in one or more of the tissues, suggesting potential novel genes. The same has been observed in other non-model organism, e.g., *Cladonia rangiferina* where only 57.2% of the assembled contigs had a BLAST hit in the nr database (Junttila and Rudd 2012; Joneson et al. 2011). Identification and analysis of pathways can help better understand the functions and interactions in symbiotic organisms such as lichens. KEGG is a knowledge base for systematic analysis of gene functions, linking genomic information with higher order functional information (Kanehisa and Goto 2000). Using Blast2GO, BlastKOALA, and GhostKOALA a total of 2,505 genes were assigned to 112 KEGG pathways. This analysis shows that the expected primary metabolic processes are active in *P. membranacea*.

The KEGG analysis associates K numbers (KO identifiers) to individual proteins in different biological pathways. Although the presence or absence of K numbers is regarded as an indication of presence or absence of certain metabolic pathways, KEGG pathway analysis

should be interpreted with caution. For e.g., no substantial evidence of a streptomycin biosynthesis pathway (Figure S29), and other antibiotic pathways (Table S17) was found and therefore the pathways were deleted. Incorrect assignments of K numbers to photosynthetic pathways were also deleted. Thus, metabolic functional associations to putative pathways and interpretations should be taken with caution since the data used here is derived from metatranscriptomics, where there can be a spillover of RNA from bacteria, cyanobionts, assorted fungi etc. during polyA capture in the library preparation.

Systems biology is a research area whose central task is to integrate many levels of omics data, from DNA to ecosystems, into predictive mathematical models that in turn can create coherent hypotheses to be experimentally tested, generating new knowledge (Garcia-Campos et al. 2015). Network analysis such as inferring networks and constructing association networks is important for biological interpretation of omics data (Fukushima and Kusano 2014). For example, a network perspective of nitrogen metabolism can be constructed for *P. membranacea* based on different transporters and differential expression of genes associated with the nitrogen model. Thus, coupling the omics data with a systems biology approach may enable the elucidation of underlying mechanisms and interconnection of the global and specific pathways present in *P. membranacea*.

2.4.5 Expression and metabolic pathway analysis

Expression analysis is the most important part of this study, where the primary objective is to investigate differences in gene expression levels in the three tissues. 12,908 (78.7%) genes were actively expressed in all the three tissues ($\log_2 \text{RPM} \geq 0.5$). To unravel the biological roles of the DEGs from the different tissues, functional annotation was carried out. GO analysis showed that the DEGs were involved in a wide variety of biological processes, such as cellular processes, binding, and metabolic processes. Certain genes and the crucial role they play in shaping the symbiosis in specific pathways in the *P. membranacea* genome are explored in detail (see below).

DEGs included the HSP proteins (HSP20, HSP30 and HSP70) that can repair denatured and misfolded proteins. HSP20 is involved in the thermal stress response, whereas, in yeast, HSP30 is induced by different stresses, including heat shock, ethanol exposure, severe osmotic stress, weak organic acid exposure and glucose limitation (Piper et al. 1997).

HSP70 is produced in response to hyperthermia, oxidative stress, and pH changes, suggesting involvement in heat stress tolerance of the mycobiont. Expression of *hsp20* and *hsp70* indicates that they are produced during seasonal changes that influence the adaptive differentiation of the *P. membranacea* local population, whereas expression of *hsp30* suggests the involvement in mycobiont heat stress tolerance. Two genes, *hsp88* and *hsp98* that are known to be upregulated at elevated temperatures (Steinhauser et al. 2016) were not differentially expressed.

The cytochrome P450 family of proteins has various oxidative functions including defense mechanisms, and stress resistance (Chapple 1998). Symbiotic-specific expression of cytochrome P450 has been seen in different symbiotic systems such as *Rhizobium* sp. BR816 in legumes (Luyten et al. 2001), and dinoflagellates in reef-building corals (Rosic et al. 2010). High expression of three cytochrome P450 genes (g9085, g11483 and g11486) in thallus suggests a potential symbiotic-specific cytochrome P450 present in *P. membranacea*.

2.4.6 Nutrient assimilation

Lichenization, i.e., mutualistic symbiosis of fungi with living algal and/or cyanobacterial cells, encompasses the organized transfer of nutrients between partners (Honegger 1996). The cyanobiont *Nostoc* sp. remains photosynthetically active and is capable of fixing both carbon (C) and nitrogen (N) (Friedl and Büdel.B 1996) which is thought to be mediated to the mycobiont as glucose (Drew and Smith 1967; Richardson et al. 1968; Hill and Smith 1972) and ammonia (Rowell et al. 1985). An overview of nutrient flow (C and N) in *P. membranacea* is illustrated in Figure 2.19.

2.4.6.1 Carbon

The mycobiont, a heterotrophic organism, derives its carbon nutrition from the photobiont (Figure 2.19). Cyanobionts provide fixed carbon to the host for biosynthesis and energy requirements. Carbon fixation takes place in the filaments of the cyanobiont (Nash 2008). In general, cyanobacteria release glucose which is rapidly and irreversibly metabolized into pyruvate via the glycolysis pathway (MacFarlane and Kershaw 1985). Transfer of nutrients, water, and metabolites between photobiont and mycobiont varies among symbiotic species (Honegger 1991). It has been hypothesized that the carbohydrate export and transfer from photobiont to mycobiont can be carried out by a polyol (ribitol,

sorbitol, and erythritol) or glucose (Collins and Farrar 1978). A ribitol transporter from *Ramalina conduplicans* (Yoshino et al. 2016) and a sugar transporter from *Cladonia grayi* (Joneson et al. 2011) have been reported. In *P. membranacea* four glucose, one sucrose, four monosaccharide, and 12 sugar transporters were identified, and some are differentially expressed in the three tissues. E.g. two MFS sugar transporters g1,183 and g1,5114 were highly expressed (> 2X FC) in apothecia and thallus, respectively. Also, it is of interest in this context that most sugar transport proteins have been lost from *Endocarpon* sp., probably during evolution (Wang et al. 2014).

Green algae are thought to release polyols (polyhydric sugar alcohols; sugar alcohol) that are taken up by the mycobionts while the carbohydrate released from the cyanobiont is glucose (Smith and Douglas 1987). The carbohydrate (such as glucose or ribitol) taken up by the mycobiont is rapidly metabolized into mannitol, via the pentose phosphate pathway (PPP) (Lines et al. 1989; MacFarlane and Kershaw 1985), and therefore made unavailable to the photobiont (Galun 1998). Accumulation of mannitol plays a major role in the stress tolerance of cyanolichens and other fungi (Ramirez et al. 2004; Drew and Smith 1967). Sugars such as glucose or ribitol are rapidly converted to mannitol, which binds additional water that can be used by both fungi and cyanobacteria under dry conditions (Ramirez et al. 2004; Drew and Smith 1967; Garg et al. 2016). In fungi and lichens such as *Pseudevernia furfuracea* and *Hypogymnia physodes*, mannitol-1-phosphate dehydrogenase (MPD; EC 1.1.1.17) catalyzes the reduction of fructose-6-phosphate to mannitol-1-phosphate using NADH as the cofactor. This is followed by dephosphorylation by a mannitol-1-phosphate phosphatase (MPP; EC 3.1.3.22) (Calmes et al. 2013; Jensen et al. 1991; (Patel and Williamson 2016) to mannitol and inorganic phosphate. Mannitol would then be oxidized to fructose by mannitol dehydrogenase (MDH; EC 1.1.1.138) using NADP⁺ as the cofactor. Finally, to complete the cycle, fructose would be phosphorylated to fructose 6-phosphate by a hexokinase (EC 2.7.1.1). In *Peltigera hymenina* mannitol is mainly localized in the middle (cyanobiont layer) and bottom layers of the lichen Garg.

In filamentous fungi, fructose-6-phosphate is converted to mannitol-1-phosphate by an MPD that is highly expressed in all mycobiont tissues in *P. membranacea* (Figure 2.19) (Meena et al. 2015). It is also possible that fructose-6-phosphate is first converted to mannitol-6-phosphate by

mannose-6 phosphate isomerase (MPI) and then to mannitol-1-phosphate catalyzed by phosphomannomutase (PMM) (Khosravi et al. 2015). Both MPI and PMM were identified in *P. membranacea* and expressed in all three tissues. Although MPP activity has been described in fungi e.g. *Sclerotinia sclerotiorum*, a search for MPP (Alternative names: Hexitol phosphatase, 2-deoxyglucose-6-phosphate phosphatase, Sorbitol-6-phosphatase, and sugar phosphatase) did not yield any sequence in GenBank for fungi, precluding a robust homology search in the *P. membranacea* genome.

Model for nutrient flow (C and N) in *P. membranacea*

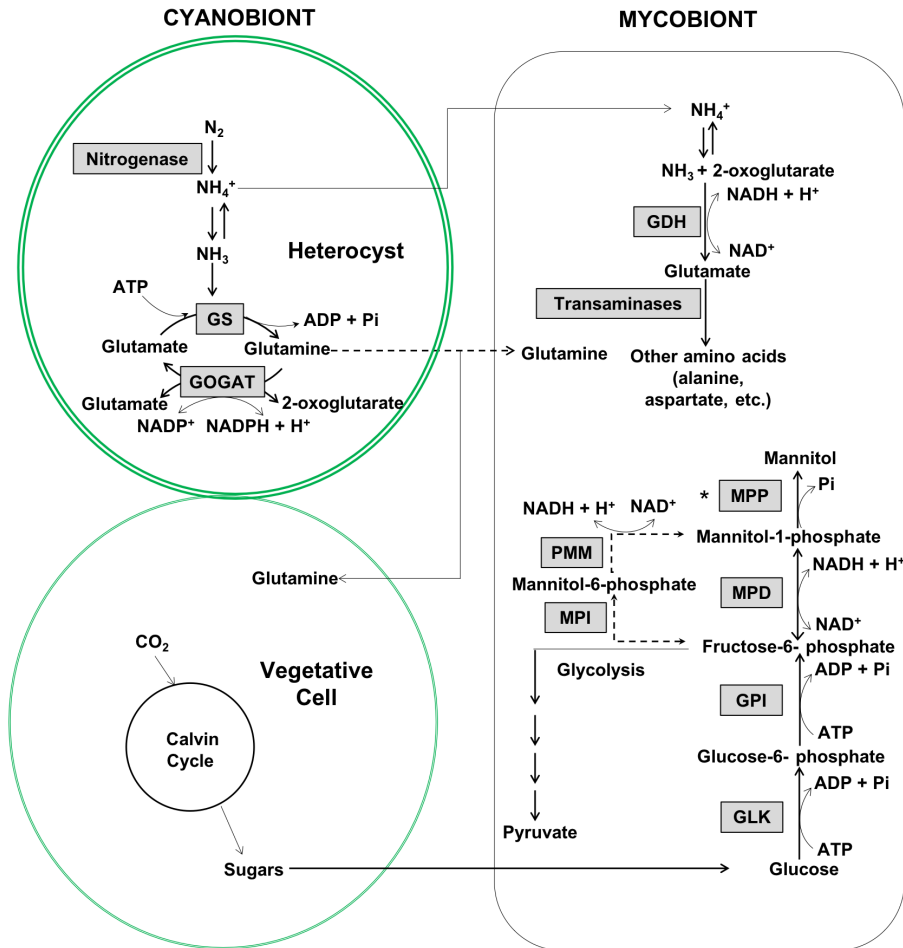


Figure 2.19 Proposed model for nutrient flow (C and N) in *P. membranacea*.

MDH, Mannitol dehydrogenase; MPP, Mannitol-1-P phosphatase; MPD, Mannitol-1-P dehydrogenase; GPI, Glucose 6-phosphate isomerase; GLK, Glucokinase; MPI, Mannose-6 phosphate isomerase alternately phosphomannose isomerase (PMI); GS, Glutamine synthetase; GOGAT, Glutamate synthase; GDH, Glutamate dehydrogenase; PMM, Phosphomannomutase. *- gene not identified in *P.mem* v1.1 genome. Dotted line- putative translocation pathway.

2.4.6.2 Nitrogen

Nitrogen assimilation in mycobionts involves uptake of either nitrate (NO_3) or ammonium (NH_4^+) ions, and possibly nitrogenous compounds such as glutamine. Ammonium ions are quite toxic (Salisbury and Ross 1992) and are rapidly converted to other nitrogen products in lichens (Rai 1988, 2002). In cyanobiont heterocysts, NH_4^+ and glutamate are joined by glutamine synthetase (GS) to form glutamine (Gln), an amide that may accumulate without being toxic and is exported to the vegetative cyanobiont cells and probably to the mycobiont. In addition, Gln and 2-oxoglutarate can form two molecules of Glu, catalyzed by GOGAT (Galagan et al. 2003). In *Peltigera* species the transferred NH_4^+ is converted to Glu by the mycobiont, presumably by GS (EC 6.3.1.2) (Stewart and Rowell 1975; Rai 2002) and two GS transcripts were identified. There was no significant difference in the GS activity between the tissues and transcripts of GOGAT, as well as both NADP^+ and NAD^+ dependent GDHs, were highly expressed in the mycobiont. The NADP^+ dependent GDH (anabolic form) catalyzes the assimilation of ammonia by reductive amination of 2-oxoglutarate to form L- glutamate, and the gene coding the enzyme was highly expressed in apothecia. NAD^+ dependent GDH which catalyzes the reverse reaction (catabolic) was relatively low in all tissues, and indicates that glutamine is being converted by GOGAT to glutamate in the mycobiont.

In *Saccharomyces cerevisiae*, *mep2*, the gene encoding the high-affinity ammonium transporter, is downregulated when NH_4^+ or Gln is high (Marini et al. 1997), therefore low expression of *mep2* in *P. membranacea* thallus suggests that the gene is subject to nitrogen repression. The low-affinity *mep3* is active during conditions when sufficient ammonium is present and it was found upregulated in the thallus of *P. membranacea*.

2.4.7 Glyoxylate pathway

Generally, glyoxylate provides a means of using acetyl-CoA to synthesize carbohydrates. The cycle plays an essential role in cell growth on acetate. Genes coding for glyoxylate enzymes were identified in the *P. membranacea* genome. One transcript of NADP^+ -dependent isocitrate dehydrogenase (*idh1*) and two transcripts of NAD^+ dependent isocitrate dehydrogenase (*idh3*) were expressed in all the three tissues of *P. membranacea*. Acetyl-CoA carboxylase (*acc1*), which catalyzes carboxylation of acetyl-CoA to malonyl-CoA was also expressed.

Consistent with the presence of the enzymes of the glyoxylate cycle, *P. membranacea* can utilize acetate and related carbon sources via the glyoxylate pathway. In the mycorrhizal ascomycete *Tuber borchii* isocitrate lyase (*icl*) and malate synthase (*mls*) were highly expressed in the fruiting bodies compared to those in the vegetative cells (Abba et al. 2007). The same was observed in *P. membranacea* where ICL, and MLS were more highly expressed in apothecia than in the rhizines.

2.4.8 Melanin pathway

Three pathways can contribute to melanin production: the DHN–melanin pathway; the DOPA–melanin pathway; and the L–tyrosine degradation pathway. In *Aspergilli*, several enzymes in these pathways are developmentally regulated as melanin is produced in conidia but not in hyphae (Youngchim et al. 2004). In *P. membranacea* homologs of the enzymes involved in the DHN–melanin pathway were identified but were expressed at low levels, especially in apothecia. The DOPA–melanin pathway genes coding for tyrosinases or laccases were also present in *P. membranacea*. In lichens, tyrosinase activity was found to be high in some taxa but low or absent in others (Beckett et al. 2012). Four tyrosinase genes were found in *P. membranacea*. One of them (g1251_scaff30955) was highly upregulated in apothecia, and the remaining three genes showed low expression in all three tissues. It is, therefore, most likely that this tyrosinase is responsible for melanin production in *P. membranacea* apothecia. An alternative pathway for melanin production, using the L–tyrosine degradation pathway, is also conserved in *P. membranacea* but most of the genes in this pathway showed very low expression.

2.4.9 Symbiotic genes

Symbiosis is of long-standing interest among lichenologists (Elifio et al. 2000; Fontaniella et al. 2004; Galun et al. 1976; Bubrick et al. 1981; Howe and Barrett 1970; Lockhart et al. 1978; Molina et al. 1998; Molina and Vincente 1995; Sacristan et al. 2006). Fungal genes implicated in the process of symbiosis by differential expression in symbiotic vs. aposymbiotic cells of *Cladonia grayi* include genes coding for proteins involved in self and non-self-recognition, lipid metabolism genes, glucose-repressible genes, an oxidoreductase, dioxygenases (Joneson et al. 2011) and probably the most important – hydrophobin (see below) (Nash 1996).

Horizontal gene transfer (HGT) in the symbionts, cyanobacteria, algae, and fungi, can make lichens better adapted over a long period. Studies have shown that mycobionts and photobionts can be recipients of genes through horizontal transfers, e.g. ammonium transporters (AMT)/methylammonium permeases (MEP) which appear to have come from hyper thermophilic chemoautolithotrophic prokaryotes (McDonald et al. 2012), and PKS genes from actinomycetes (Schmitt and Lumbsch 2009). Similarly HGT has been identified in other organisms, e.g. proline racemase from bacteria to *Candida parapsilosis* (Fitzpatrick et al. 2008), and MAPK1 (mitogen-activated protein kinase-1) from bacteria to *Metarhizium anisopliae* (Duan et al. 2009).

AMT/MEP proteins are found throughout all domains of life and are crucial for the uptake and translocation of NH_4^+ across the cell membrane (Tremblay and Hallenbeck 2009; Andrade and Einsle 2007; Ninnemann et al. 1994). In addition to the above, under nutrient starvation (lack of N or C), a mycorrhizal MEP acts as an ammonium sensor that is essential in switching between two morphological fungal growth forms, a cellular yeast form and a multicellular invasive filamentous form (Javelle et al. 2003). High-affinity ammonium transporters were highly expressed in tissues where there are no cyanobacteria (e.g. in rhizines and apothecia) whereas low-affinity ammonium transporter was highly expressed in the thallus where cyanobacteria reside near mycobionts.

Hydantoin racemase, an enzyme that catalyzes the racemization of 5-substituted hydantoin to produce L-amino acids (Martinez-Rodriguez et al. 2004) is present in *Sinorhizobium meliloti*, a nitrogen-fixing bacterium that forms a symbiotic relationship with legumes. Hydantoin racemase is expressed highly in thallus tissue, suggesting it could be important for symbiosis. MAPKs, serine/threonine-specific protein kinases, plays diverse roles in the regulation of gene expression, cellular growth, and survival (Pearson et al. 2001). A MAPK acts as a signaling mediator in the symbiotic interaction between soybean and *Bradyrhizobium japonicum* (Fernandez-Pascual et al. 2006). Four *mapk1* genes were identified, and g14647 (*spm1*) was differentially expressed in *P. membranacea*, suggesting a possible role in symbiosis.

Hydrophobins have a key role in making the hyphal wall surfaces water repellent. It prevents waterlogging of the fungal surfaces that are in contact with air thereby isolating the photobiont and making the

photobiont dependent on mycobiont for water and nutrients (Talbot et al. 1996; Lugones et al. 1996; Bell-Pedersen et al. 1992; Wosten et al. 1994). In fungi, they are secreted during various processes such as sporulation, fruiting body development and infection structure formation (Kershaw and Talbot 1998). In lichens such as *Xanthoria* spp., hydrophobins hold the mycobiont and photobiont together (Honegger 1991; Wessels 1997). In *P. membranacea*, a gene encoding hydrophobin H1 was identified and was highly expressed in the thallus tissue.

In the lichen-forming ascomycetes, lectins are thought to be potential mediators of symbiosis involved in photobiont partner recognition (Elifio et al. 2000; Fontaniella et al. 2004; Galun et al. 1976; Bubrick et al. 1981; Howe and Barrett 1970; Lockhart et al. 1978; Molina et al. 1998; Molina and Vincente 1995; Sacristan et al. 2006). Two genes encoding galectin-like lectins in *P. membranacea*, *lec-1* and *lec-2* (Miao et al. 2012) have been identified and characterized. *lec-1* (168 aa) was identified using two PMA (*P. membranacea* agglutinin) peptides (Lehr et al. 1995) as a query and *lec-2* (159 aa) was identified using *lec-1* as a query during BLAST searches. Five transcripts of *lec-1* (Figure S10) and one transcript of *lec-2* (Figure S11) have been identified in P.mem v1.1. *Lec-1* and *lec-2* were highly expressed in thallus and apothecia, respectively. *Lec-1* is implicated in establishment or maintenance of symbiosis, and although *lec-2* shows a high degree of polymorphism, it does not appear to be a determinant of photobiont selection (Manoharan et al. 2012; Miao et al. 2012). The *lec-1* gene was not affected significantly at increased temperatures whereas *lec-2* was downregulated at increased temperatures (Steinhauser et al. 2016).

Fucose binding lectins specifically recognize fucosylated glycans, e.g., AAL (*Aleuria aurantia* lectin). They have been identified in the lichen *Endocarpon pusillum* (Wang et al. 2014) and in the fungus *A. fumigatus* where AFL (*Aspergillus fumigatus* lectin) is thought to be an important virulence factor involved in an early stage of *A. fumigatus* infection (Houser et al. 2013). Four fucose binding lectins were highly expressed in mycobiont tissues with two genes more highly expressed (>2X fold) in rhizine than apothecial tissue. The legume lectin Concanavalin A (*ConA*) from *Canavalia ensiformis* binds to glucose/mannose residues (Goldstein et al. 1965) and two *ConA* homolog genes were identified and highly expressed in thallus tissue in *P. membranacea*.

2.4.10 Mating-type and heterokaryon incompatibility genes

Sexual reproduction in ascomycota is controlled by genes encoded at the mating type (*mat1*) locus. In heterothallic species, the two allelic versions, or idiomorphs of the mating locus are *mat1-1-1* (for the *mat1-1* idiomorph) and *mat1-2-1* (for the *mat1-2* idiomorph) genes (Wilken et al. 2018). When both idiomorphs, *mat1-1* and *mat1-2* are expressed in a single individual this indicates that the fungus is homothallic (Wilson et al. 2015). Putative *mat* alpha-box (*mat1-1-1*) and HMG-box (*mat1-2-1*) domains were found in P.mem v1.1, but the *mat* genes were present in two different scaffolds and it was difficult to reveal the normal chromosomal position relative to each other, or to *apn2* or *sla2*, the two conserved genes that typically flank *mat1* in ascomycetes. P.mem v1.1 was generated from mixed thalli and it was not possible to determine if the mating type genes would have been together in a single genome (homothallic) or separate (heterothallic). However, both the idiomorphs were identified in the sequences generated from a single lobe (A. Gagunashvili, unpublished) suggesting that *P. membranacea* is homothallic.

Genes involved in fruiting body development and sexual ascus development (rhamnogalacturonase B) were also in present. Their orthologs in *N. crassa* and *A. nidulans* are known to be required for meiosis and sexual development. Homothallic fungi, e.g., *C. albicans* can also engage in parasexuality under suitable circumstances (Bennett & Johnson 2003). The gene families involved in the parasexual cycle in fungi are heterokaryon self-incompatibility (Fischer et al. 2009) genes involved in the initiation process, the self/non-self-recognition genes such as heterokaryon incompatibility (Sheth et al. 2006) or vegetative incompatibility (*vic*) genes involved in the formation of interhyphal anastomoses, modifier (*mod*) and suppressor (*sup*) genes that are involved in maintenance and signaling, and programmed cell death genes (*pcd*) involved in apoptosis (Pal et al. 2007). Expression of heterokaryon incompatibility related genes, including *mod* and *sup* genes, implies that a parasexual life cycle is present in *P. membranacea*. It also has a complement of genes for conidiophore development, and conidiophores are found in Peltigerales, e.g. *Lobaria laetevirens* (Honegger 1984) although they have not been reported for *P. membranacea* (Table 2.10).

2.4.11 Caveats to the study

There were limitations in the data and technology at the time of the study. Also, at the time of sampling the technology was rapidly developing in terms of read length and total data output. The transcriptomic data was generated from short read technologies (36-110 bp), and in some cases it was difficult to predict the whole gene set in a single scaffold. The transcriptomic data generated for *P. membranaceae* showed uneven coverage. When the annotated genes and P.mem v1.1 genome were compared to the transcripts from the four *de novo* transcriptomes, 561 *de novo* transcripts were not represented in the genome possibly due to misassembly of P.mem v1.1 or due to the overrepresentation of transcripts as the *de novo* transcripts were assembled for each tissue and combined later. Also, the cost of the sequencing played an important role in determining the output of the data generated per sample which led to a restriction in the read coverage and in the number of samples to be sequenced for the study. Also, some biological limitations were incurred, for instance, not all three tissues were sampled during all seasons because of the lack of apothecia in some years and seasons, e.g. June and July. The DEGs from the nine comparisons included many hypothetical proteins, uncharacterised and predicted proteins. Many of the genes associated with tissue-specific expression had no known function and therefore might be novel and need investigation in the future. Because of the limitations mentioned above, we decided to choose the gene-based pathway analysis along with the transcriptomic data for the global metabolic pathways. Differential expression of genes identified in this study for certain pathways is not statistically significant. However, function of such genes with similar expression profiles have been shown in other studies.

The *P. membranacea* draft genome assembly consists of 3,300 scaffolds with gaps and it appears that the unbridged gaps are primarily due to a fairly high number of repeat elements (3.97%). Nevertheless, we strongly believe that this study and the data generated will serve as a good source for future studies not only for *P. membranacea* but lichens in general. The genome assembly includes 96.9% of the conserved BUSCO core gene set. With recent advances in sequencing technology such as long read technologies (>20 Kb), e.g. single molecule real-time (SMRT) sequencing by Pacific Biosciences (PacBio) (Eid et al. 2009) and nanopore sequencing by Oxford Nanopore Technologies (Clarke et al. 2009), there will be opportunities for hybrid analysis with existing data

which should resolve the shortcomings of the sequencing technologies used and help improve the genome assembly of *P. membranacea* in the future. Of note, single-molecule direct RNA sequencing without cDNA synthesis (e.g., that of Helicos BioSciences Corporation) should reduce biases in the quantification of transcripts (Ozsolak and Milos 2011).

2.5 Conclusion

We report here the first tissue-specific transcriptomic profiling of a cyanolichen with implications for other lichens species and fungal tissue differentiation. We also present the first draft assembly of *P. membranacea* (P.mem v1.1) with annotation guided by transcriptomic data from different tissues collected under diverse environmental conditions. The analysis presented here provides significant information on the biological functions within each of the tissues and provides insights into the differential regulation of genes and pathways in each of the tissues. Therefore, the study will provide a useful resource for the lichen research community for further studies on genomics, gene expression, and functional studies in *P. membranacea* and the results can be considered a benchmark for further investigation in cyanolichens in general.

CHAPTER 3

DNA CYTOSINE METHYLATION AND GENOME DEFENSE IN THE LICHEN *PELTIGERA* *MEMBRANACEA*

3 DNA CYTOSINE METHYLATION AND GENOME DEFENSE IN THE LICHEN *PELTIGERA MEMBRANACEA*

3.1 Introduction

Epigenetic modifications are heritable changes in gene activity and expression that occur without alteration in DNA sequence (Bird 2007; Goldberg et al. 2007). In eukaryotes, the chromatin states and genome activity are regulated by a variety of different epigenetic mechanisms, mainly DNA methylation, RNA interference signals (RNAi-mediated silencing) and post-translational modification of histones (e.g. methylation, acetylation, and phosphorylation). DNA methylation is the best studied and most conserved epigenetic modification.

DNA methylation occurs at adenine as well as cytosine bases. In prokaryotes, adenine (N-6 position) or cytosine methylation (C-5 or N-4 positions) plays a central role in the host restriction/modification system (Arber 1974; Boyer 1964; Roberts 1978). It has been widely accepted that DNA methylation in eukaryotes is generally restricted to cytosine bases (C-5 position) and exists in most eukaryotic organisms including fungi, protists, flowering plants, all vertebrates, some invertebrate insects and to a lesser extent in certain bacterial species (Chan et al. 2005; Law and Jacobsen 2010). However, adenine methylation (N-6 position) is present at low levels in plants and animals (Wu et al. 2016; Zhang et al. 2015; Fu et al. 2015). In fungi, cytosine methylation suppress the expression of transposable elements and is present in gene bodies that have been shown to be involved in modulating transcriptional activity e.g. in *A. immersus*, *Candida albicans* and *Uncinocarpus reesii* (Barry et al. 1993; Zemach et al. 2010; Mishra et al. 2011; Seidl 2017). Whereas, noncanonical N6-methyldeoxyadenine (6mA) methylation is associated with transcriptionally active genes in early-diverging fungal lineages but is not detectable in ascomycetes (Mondo et al. 2017).

Cytosine DNA methylation involves the addition of a methyl group to the fifth carbon of the cytosine pyrimidine ring to produce 5-methylcytosine (5mC) with S-adenosylmethionine (SAM) serving as the methyl donor (Holliday and Pugh 1975; Mato et al. 1997; Riggs 1975). Cytosine DNA methylation plays an essential role in the normal function and development of multicellular eukaryotes and is associated with various biological processes, such as silencing of transposons, regulation of gene expression during development and cell differentiation (Jones and Taylor 1980; Law and Jacobsen 2010). Ten-eleven translocation proteins (TET1, TET2, and TET3) which oxidize 5mC to 5hmC (5-hydroxy methyl cytosine), an intermediate of 5mC demethylation, have been implicated in epigenetic modifications that influence chromatin structure (Tan and Shi 2012).

Cytosine DNA methylation is established and maintained by a family of conserved DNA methyltransferases (DMTs). Eukaryotic DMTs have been grouped into six subfamilies, namely DNMT1/MET1, DNMT2/ PMT1, DNMT3/DRM, a chromomethylase (CMT), MASC1/RID (RIP defective), and MASC2/DIM2 (Chen and Li 2004). The DMTs are subdivided into 'maintenance' and '*de novo*', depending on whether the recognition site is already methylated or not (Riggs 1975). Fungal DMTs such as *Neurospora* DIM-2 and *Ascobolus* MASC-2 belong to the 'maintenance' methyltransferase family (Kouzminova and Selker 2001; Malagnac et al. 1997; Goll and Bestor 2005). In maintenance methylation, the methyltransferase copies the DNA methylation pattern from the parental strand to the daughter strand during semi-conservative DNA replication (Holliday and Pugh 1975; Riggs 1975). In *de novo* methylation, DMTs catalyze the transfer of a methyl group to unmethylated DNA (at symmetrical or asymmetrical sites) to establish a new methylation pattern, e.g. with DNMT3a and 3b (Okano et al. 1999).

Dim-2 (defective in methylation) (Klocko et al. 2015) encodes a DMT responsible for all cytosine DNA methylation in vegetative tissues of *Neurospora crassa*, i.e. both maintenance and *de novo* methylation (Foss et al. 1993; Kouzminova and Selker 2001). DIM-2 methylates repeat elements (such as transposons), and DNA methylation depends on methylation of lysine 9 of histone H3 (H3K9) (Kouzminova and Selker 2001; Tamaru and Selker 2001; Tamaru et al. 2003). H3K9 is associated with gene silencing in many animals, plants, and fungi (Lewis et al. 2010).

In *N. crassa*, DCDC, a five-protein complex (DIM-5, DIM-7, DIM-9, CUL-4, and DIM-8 (DDB-1)) is essential for trimethylation of lysine 9 on histone H3 (H3K9me3) and DNA methylation (Zhao et al. 2010; Lewis et al. 2010b; Lewis et al. 2010a). DIM-7 is required for DIM-5 localization and mediates the interaction of DIM-5 with DDB1 (DNA-damage-binding protein 1)/CUL4 through DIM-9 (or DDB-1/CUL-4 Associated Factor, DCAF) (Figure 3.1) (Lewis et al. 2010b; Xu et al. 2010). DIM-5 is sensitive to modifications of histones (e.g. methylation of Lysine 4 and phosphorylation of Serine 10 of histone H3) and specifically methylates histone H3K9. The H3K9me3 mark is recognized by heterochromatin protein 1 (HP1) through its chromodomain (Freitag et al. 2004). HP1 recruits DIM-2, a histone demethylase complex (DNA methylation modulator, DMM), and a histone deacetylase complex (HCHC) (Honda et al. 2010; Honda and Selker 2008; Honda et al. 2012). The jumonji domain protein complex, DMM (comprised of DMM-1, DMM-2 and HP1), limits spread of heterochromatin and subsequent DNA methylation from AT-rich regions in genes (Honda et al. 2010). HCHC consists of HP1, chromodomain protein 2 (CDP-2), the histone deacetylase HDA-1, and the AT-hook motif protein CDP-2/HDA-1-associated protein (CHAP) and is required for proper chromosome segregation (Honda et al. 2016).

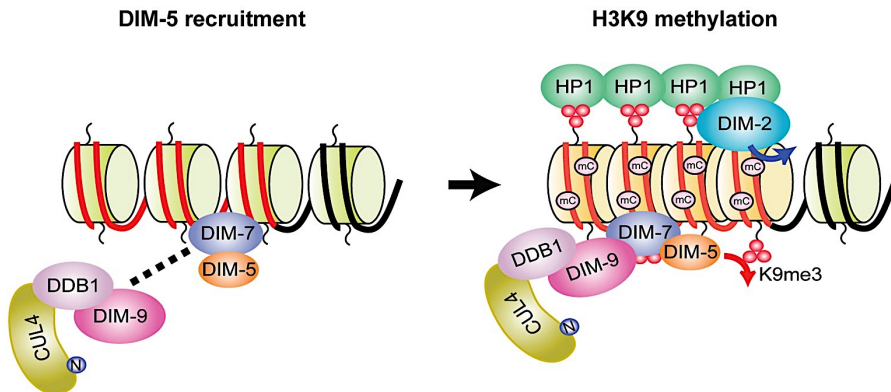


Figure 3.1 Heterochromatin formation and DNA methylation in *N. crassa*.

(See text for details. Figure reproduced from Lewis, 2010).

Fungal-specific DMTs such as RID (RIP defective) (Freitag et al. 2002) and MASC1 (Malagnac et al. 1997; Irelan and Selker 1996) are associated with RIP (Repeat-Induced Point Mutation) in *N. crassa* and

MIP (Methylation Induced Premeiotically) in *Ascobolus immersus* and *Coprinus cinereus*, respectively. Unlike MASC1, which is also involved in sexual development, RID is not required for vegetative or sexual development (Freitag et al. 2002; Malagnac et al. 1997). In *Neurospora*, methylated DNA is mainly found in transposable elements (TE) and repetitive elements that have undergone RIP. In this process, duplicate sequences detected during the dikaryotic stage of the *N. crassa* sexual cycle are silenced permanently. RIP, the first eukaryotic genome defense system discovered, introduces C to T (or G to A) transition mutations that often led to nonsense or missense mutations that affected the expression of the sequences (Hane et al. 2015; Cambareri et al. 1989). In *N. crassa*, RIP mutations only accumulate within repeated sequences that are at least 400 bp in length (Watters et al. 1999). RIP mutated sequences are frequent (although not always) targets for *de novo* DNA methylation in vegetative cells (Freitag et al. 2002; Selker 1990; Galagan and Selker 2004; Rosa et al. 2004; Mautino and Rosa 1998). It is likely that the RIP mutated bases are derived by enzymatic deamination of 5mC or by deamination of Cs followed by DNA replication (Selker 1990). This was proposed because cytosine methylation involves an intermediate that is prone to deamination catalyzed by a RID (Freitag et al. 2002). A similar process, MIP in *A. immersus* and *C. cinereus* occurs before the sexual cycle, i.e. during the period between fertilization and karyogamy. MIP employs DNA methylation to reversibly silence repeated sequences (Malagnac et al. 1997; Goyon and Faugeron 1989; Rhounim et al. 1992; Barry et al. 1993).

Methylation patterns differ by organisms. In animals, DNA methylation is restricted to the symmetric (both strands) CG context. In plants, cytosine methylation occurs symmetrically in both CG and CHG (H=A, T, or C) contexts, and also asymmetrically in the CHH context, with the latter directed and maintained by small RNAs (Law and Jacobsen 2010). In fungi, e.g. *N. crassa*, cytosine methylation occurs at both symmetrical and asymmetrical sites and all sequence contexts: CG, CHG, and CHH (Suzuki and Bird 2008).

DNA methylation in fungi has primarily been viewed as a mechanism of genome defense to silence TE and other repeats (Giannino et al. 2003). The ascomycete *N. crassa* shows substantial DNA methylation, approximately 1.5% of the total cytosines are methylated in vegetative cells (Foss et al. 1993). In contrast *Saccharomyces cerevisiae* (Proffitt et al. 1984), *Schizosaccharomyces pombe* (Antequera et al. 1984;

Wilkinson et al. 1995) and *Aspergillus nidulans* lack DNA methylation (Lee et al. 2008). *Aspergillus flavus* shows very little cytosine methylation (0.04%), while the ascomycete *Tuber melanosporum* (pezizomycetes), with high repetitive DNA content (>58%), has the highest level of methylation with >44% of mCs sites methylated primarily in repeated sequences (Gowher et al. 2001; Liu et al. 2012).

Homology-dependent gene silencing (HDGS) can be divided into transcriptional gene silencing (TGS) where the transcript level is limited by suppressing the transcription, and post-transcriptional gene silencing (PTGS)/RNA interference (RNAi) where a sequence-specific RNA degradation process is activated (Cerutti and Casas-Mollano 2006; Catalanotto et al. 2002). In filamentous fungi, duplicated sequences are silenced by the following mechanisms. So far in filamentous ascomycetes, TGS mechanisms such as RIP and MIP (see above), PTGS mechanisms such as quelling (Cosuppression) (Cogoni and Macino 1997, 1999; Romano and Macino 1992) and meiotic silencing of unpaired DNA (MSUD) (Shiu et al. 2001) have been reported.

PTGS is part of a broad host defense response that controls invading transposons and provides protection of the genome in most eukaryotic organisms, including filamentous fungi (Cogoni and Macino 1999; Hammond et al. 2008). In quelling, small interfering RNAs (siRNAs) are produced specifically from chromosomal loci that contain repetitive DNA (Chang et al. 2012). In MSUD chromosomal segments of “unpairable” DNA become reversibly silenced via a mechanism similar to quelling but only present during meiosis (Shiu et al. 2001). Proteins involved in PTGS/RNAi are present in all lineages of eukaryotes although the numbers of genes involved in the RNAi machinery vary. In general, filamentous ascomycetes have three core RNAi components, RNA-dependent RNA polymerase (RdRP), Dicer, and Argonaute (Cerutti and Casas-Mollano 2006; Shabalina and Koonin 2008). In this pathway, an aberrant RNA (aRNA) is first recognized and converted into double strand RNA (dsRNA) by RdRPs, subsequently Dicers cleave the dsRNA into siRNA duplexes (~25 nt) that are loaded onto Argonaute proteins, resulting in the activation of an RNA-induced silencing complex (RISC) (Nakayashiki et al. 2006). Cellular components of RNA silencing in *N. crassa* include the RNA-dependent RNA polymerases QDE-1 and SAD-1, the Argonaute-2 orthologs QDE-2 and SMS-2, a RecQ helicase, QDE-3, and two Dicer orthologs, DCL-1 and DCL-2 (Catalanotto et al.

2004; Catalanotto et al. 2002; Cogoni and Macino 1999, 1997; Lee et al. 2003; Shiu et al. 2001) (Figure 3.2).

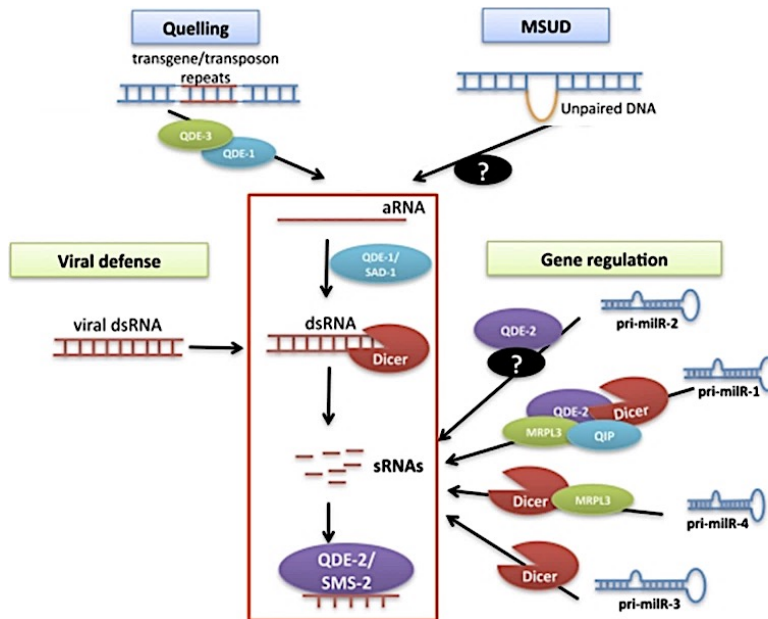


Figure 3.2 sRNA biogenesis pathways and RNAi functions in *Neurospora*.

(See text for details. Figure reproduced from Dang et al. (2011)).

Till date, only two studies on DNA methylation in lichens have been reported. Both studies employ green algal lichens (chlorolichens). In the first study, restriction analysis of DNA from *Cladonia grayi* using CpG methylation-sensitive restriction endonucleases showed a correlation between symbiosis and methylation of fungal DNA in *C. grayi*. The authors concluded that overall DNA methylation is high in symbiotic tissue (podetia) and vegetative propagules (soredia), but low in the fungal ascocarps (apothecia) and cultured mycobiont (Armaleo and Miao 1999). In the second study, DNA methylation in *C. grayi* was determined using bisulfite sequencing where the symbiotic structures such as podetia, squamules, soredia and the aposymbiotic cultured mycobiont and alga were sequenced. The study reported that there is no differential DNA methylation in the genome of the mycobiont in the symbiotic and aposymbiotic states, whereas the genome of the alga is methylated both in the symbiotic and aposymbiotic states (McDonald 2011). The results indicated that there is no correlation between DNA methylation and symbiosis.

While the tissues from the chlorolichen *C. grayi* have been subjected to BS sequencing (McDonald 2011), the method has hitherto not been applied to a cyanolichen. The present study was undertaken to study DNA methylation of a cyanolichen and more specifically to examine the DNA methylation across the tissues, as methylation levels and patterns may differ between different tissues. The objectives of this study were to 1) generate and characterize a lichen methylome using next-generation sequencing technology, 2) to gain information on tissue-level variation in methylation, and 3) to identify putative genes involved in DNA methylation and RNA silencing related-phenomena. In this study, the methylomes from thallus and rhizines of the cyanolichen *P. membranacea* were determined to the single base level using bisulfite (BS) sequencing. This is the first time DNA methylation has been characterized in *Peltigera membranacea*. The data will serve as a useful resource to better understand the epigenesis and functional genomics in cyanolichens.

3.2 Materials and Methods

3.2.1 Lichen Material, DNA and RNA isolation

Apothecia (A), Thallus Whole (TW) from autumn 2009 and rhizines (R), thallus body (TB-thallus without rhizines), thallus whole (TW-thallus with rhizines) from summer 2012 were collected from Keldur; Reykjavik, Iceland. Samples were collected in RNeasy lysis buffer and transported to the laboratory. Rhizines were separated from the thallus body while allowing the tissue to be immersed in RNeasy lysis buffer. Each thallus lobe was cut in half. One part of the lobe was left in RNeasy lysis buffer for RNA isolation while the other half was washed briefly in water, cleaned, air-dried and stored for genomic DNA isolation. Total RNA was isolated using the RNeasy RNATM kit and was isolated within 24 hours of collection to minimize RNA degradation. The isolated RNA was run on a 1% agarose denaturing gel. Genomic DNA was extracted by using methods described previously (see Appendix A2) (Xavier et al. 2012) and run on a 1% agarose gel. Before library preparation, each sample was subjected to quality control as follows. The concentration and purity of RNA and DNA were assessed by a NanoDrop spectrophotometer (NanoDrop® ND1000; Thermo Fisher Scientific Inc., Waltham, MA, USA). For RNA, the sample was required to have an A260/A280 ratio between 2.0 and 2.2 and for DNA between 1.8 and 2.0 with an A260/A230 ratio above 2.0. RNA integrity was also

further evaluated using the RNA Nano Chip run on an Agilent 2100 Bioanalyzer (Agilent Technology Inc.) (Figure S5).

3.2.2 Background of *P. membranacea* genome

Previously, the lichen *P. membranacea* was sequenced at the University of Iceland using both 454 and Illumina technologies (Xavier et al. 2012). The *P. membranacea* 454 DNA contigs (Pmem 454Allcontigs) consist of sequences from the photobiont *Nostoc*, the mycobiont, and assorted bacteria. P.mem v1.0 consists of only mycobiont contigs (14,674) which were assembled into scaffolds (3,033). P.mem v1.1 is the annotated version of P.mem v1.0. The statistics of *P. membranacea* 454allcontigs and P.mem v1.1 genome assembly are provided in Appendix A3. The repetitive sequences in the *P. membranacea* genome (P.mem454 AllContigs and P.mem v1.1) were masked using Repeatmasker v3.3.0 (<http://www.repeatmasker.org>) with default parameters (Table S29).

3.2.3 Bisulfite treated whole genome DNA sequencing (WGBS)

One microgram of genomic DNA from two biological replicates for each of the three tissue types (summer R, TB, and TW from 2012) were treated with sodium bisulfite following the instructions included with the EpiTect Bisulfite Kit (Qiagen). Untreated genomic DNA was used as a control. Library construction and sequencing of the bisulfite treated DNA and control DNA were performed exactly as described in the Illumina Whole-Genome Bisulfite Sequencing for Methylation Analysis protocol and TruSeq® DNA Sample Preparation Guide, respectively. In brief, for the construction of bisulfite libraries, DNA was fragmented and after blunting ends and 3'-end addition of dA, Illumina methylated adapters were ligated followed by bisulfite treatment and amplification by PCR (Figure 3.3). One and two libraries were generated for each control and case samples, respectively. The libraries generated were sequenced using an Illumina Genome Analyzer at commercial facilities (deCODE Genetics), generating paired-end (PE) 101 base reads.

3.2.4 BS-Seq data processing

Raw Bisulfite-Seq reads were trimmed to remove both low-quality reads and adapters using Trim Galore (v0.3.3, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters. Duplicate reads were removed using the CLC Genomics Workbench. FastQC (v0.10.1, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was performed on processed BS-Seq reads to check the quality of the reads

before and after trimming. For the alignment of processed bisulfite reads different programs including BRAT (Harris et al. 2010), BSMAP (Xi and Li 2009), Bismark (Krueger and Andrews 2011) and CLC Genomics Workbench were used on a test dataset (data not shown). CLC Genomics Workbench achieved a higher mapping efficiency than Bismark, a widely used tool.

A key challenge with whole-genome bisulfite sequencing is the accurate alignment of the bisulfite sequences to the reference genome. The challenge arises because any given C in the reference genome can correspond to either C or T in bisulfite sequencing libraries, depending on its methylation status. mC is identified as C in sequencing reads aligned to genomic C, while unmethylated C is identified as T. On the complementary strand mC is identified as G in sequencing reads aligned to genomic G, while A is identified as unmethylated. In the CLC Genomics Workbench: 1) the reads were converted *in silico*, i.e. all C in read 1 of a pair is converted to T, and all G in read 2 of a pair is converted to A (complementary to C–T conversion) and 2) the reference genome was converted into two versions, i.e. C to T or G to A reference genome versions. A non-directional mapping was performed thereby producing alignments to four strands; original top strand (OT), original bottom strand (OB), complementary to OT (CTOT) and complementary to the OB (CTOB). Therefore, two different conversions of the read pair were created, i.e. the first conversion consists of CT conversion of read 1 and GA conversion of read 2 and second conversion consists of GA conversion of read 1 and CT conversion of read 2 (Figure 3.3). Then the *in silico* converted processed reads from case (BS treated) and control samples (Table S33) were mapped as single reads to the two conversions of *P. membranacea* scaffolds (P.mem v1.1) and mitochondrial (mt) genome (accession no: NC_016957) using CLC Genomics Workbench v9.0.1. The following parameters were used for mapping: mismatches–3; insertions–3; deletions–3; length fraction–1; similarity fraction–0.9, and for each read pair the best of the four mapping options is reported as the final result (Table S33). For each tissue, the technical replicates were combined for better coverage (Table 3.1, Table S32) (Raine et al. 2018). A methylation profile of individual cytosines for the whole genome, exons, and introns was generated for all contexts (CG, CHG, and CHH) using the 'call methylation levels' tool in CLC Genomics Workbench. The methylation levels were compared using Fisher's Exact test with a 5% FDR correction in 1000 bp consecutive windows containing at least five differentially methylated cytosines (mCs) to find the differentially

methyated regions in each tissue. A methylation bias plot was generated for each sample and based on the plots, the first eight bases from the 5' end were ignored/clipped by checking the 'read soft clip' option to avoid biases in methylation calling.

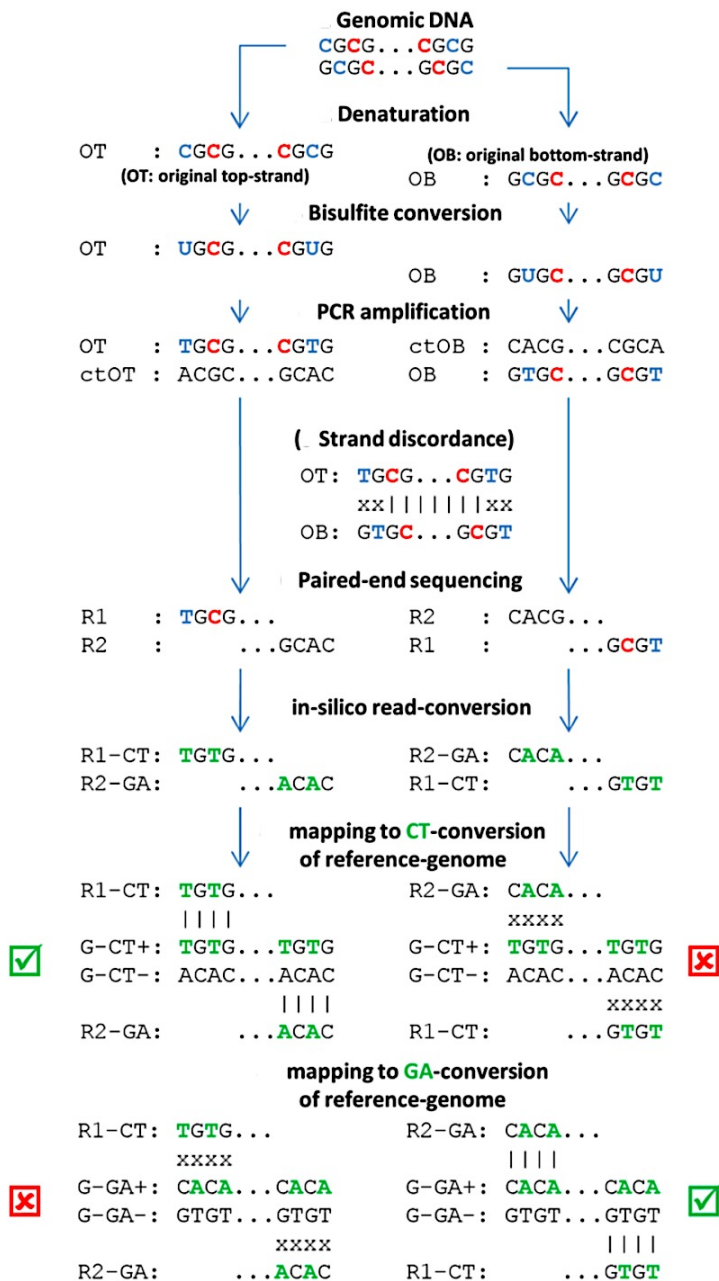


Figure 3.3 BS-Seq workflow, from generation of reads to mapping of the reads.

(Figure reproduced from Kelly et al. (2012)).

3.2.5 Transcriptome sequencing and analysis

mRNA was purified from 8–9 µg of total RNA from A (autumn-2009), R and TB (Summer 2012) samples. Multiplex sequencing was done on the tissues using the TruSeq™ RNA sample preparation protocol from Illumina. The mRNA libraries were sequenced using the Illumina Genome Analyzer at commercial facilities, generating paired-end (PE) 100 base reads.

For the expression analysis, a reference-based approach using TopHat (Trapnell et al. 2009) was carried out. Firstly, the processed RNA-Seq FASTQ reads from the above tissues were mapped to the P.mem v1.1 reference genome using Bowtie2 (v2.1.0.0) (Langmead and Salzberg 2012), and TopHat (v2.0.8b) (Trapnell et al. 2009), using default parameters. The output from TopHat, i.e. aligned reads in BAM format was viewed and analyzed using Seqmonk (<http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/v1.37.1>). Pseudo chromosomes were created using the build custom genome option implemented in Seqmonk and features such as mRNA, CDS, exons, and introns were added using the gff annotation file of the *P. membranacea* genome (P.mem v1.1). Data sets and replicate sets were defined accordingly using the 'edit data sets' and 'edit replicate sets', respectively present in the data menu in Seqmonk. For the transcriptome analysis, mRNA was quantified using the RNA-Seq quantitation pipeline on merged transcripts isoforms counting reads over each mRNA feature quantified in log₂ RPM (reads per feature per million reads of (nonstrand-specific library) and in raw read counts (Wan et al. 2016). The cumulative distribution for the data was adjusted across the sample at the 75th percentile (Langmead et al. 2010). Lists of differentially expressed genes (DEGs) were generated from the normalized quantified data using an intensity difference filter (within Seqmonk) with multiple testing corrections, and an FDR correction at $p < 0.05$ (Brien et al. 2013). An intensity difference test is a pairwise statistical test used to compare two or more data. First, a local distribution of the differences for probes (mRNA) with similar average intensity to the probe (mRNA) being examined is constructed. The set of differences between the two data sets is then modeled, i.e. the likelihood of whether a data to be an outlier when compared with the overall normal distribution of the data is examined. P value is calculated for the changes in variability at high and low intensities. In general, highly expressed genes have low variability, and genes with low expression have high variability. For each of the tissue comparisons, a pairwise comparison was made, i.e. A vs. R, A vs. TB, and R vs. TB. The list of genes from each comparison was

collated using the 'collate existing lists' option available in the filtering menu in Seqmonk.

3.2.6 Southern blotting

Genomic DNA was extracted (as above) from thallus and apothecia of *P. membranacea* collected in autumn 2009. DNA from a *Nostoc* sp. cultured from *P. membranacea* was used as a negative control since it had no homology to the probe regions. DNAs were digested using the isoschizomers *DpnII* (cleaves GAT↓C) and *Sau3AI* (blocked by GAT↓C hemimethylation). Gel electrophoresis, blotting, probe labeling, hybridization, and washing were performed as described in Appendix I. DNA probes were synthesized by PCR amplification of the genomic DNA from a repeat element, gypsy transposon (1648 bp) and the first wholly annotated gene from *P. membranacea*, DEAD DEAH box RNA helicase gene (*ded-1*) (550 bp). The probes were labeled using α -³²P dCTP. Digested DNA was electrophoresed and transferred from the agarose gel to Amersham Hybond-N+ nitrocellulose membrane (GE Healthcare Life Sciences) and bound labeled probe was detected using a Typhoon gel and blot imager (9410, GE Healthcare Life Sciences).

3.2.7 Identification and manual annotation of genes

The genes involved in the DNA methylation machinery and the RNA interference pathway were identified and annotated from the *P. membranacea* genome using the homologous genes from *N. crassa* as queries. A search was also done to identify and annotate the RID (RIP defective) protein in *P. membranacea* as well as from eight chlorolichen species sequenced at Duke University (Daniele Armaleo, personal communication, January 16th, 2012) (Table S31), using the homologs from *N. crassa* (accession AF500232) and *Ascobolus immersus* (*mascl* [*rid* homolog], accession AF025475).

3.2.8 RIP analysis of *P. membranacea* repeat families

The TE sequences of *P. membranacea* were identified using <http://repeatmasker.org>) which uses the TE library Repbase (<http://www.girinst.org/replib/>). The TE sequences were aligned using CLC Genomics Workbench. The dinucleotide frequencies were determined using RIPCAL (Oliver and Hane 2008) and RIP indices (TpA/ApT and (CpA+TpG)/ (ApC+GpT) for TEs and whole genome sequences (Pmem45Allcontigs, P.mem V1.0, and P.mem v1.1) were subsequently calculated. The TpA/ ApT index measures the frequency of RIP products (TpA) and corrects for false positives that arise from ApT-rich regions.

High TpA/ApT values indicate a strong RIP response. The (CpA + TpG)/ (ApC + GpT) index measures the depletion of the RIP targets CpA and TpG. In this index, a low (CpA+ TpG)/ (ApC+GpT) score strongly suggests RIP. The standard reference values of the RIP indices are: TpA/ApT ratios >0.89 and/or (CpA+TpG)/ (ApC+GpT) ratios <1.03 (Oliver and Hane 2008).

3.3 Results

3.3.1 DNA methylome of *P. membranacea* tissues

Whole genome DNA methylation profiles were generated using BS-Seq for the following *P. membranacea* tissues: rhizines (R), thallus body without rhizines (TB), and thallus whole (TW) using BS-Seq. BS-Seq of the samples ranged from 27–31 million reads, and a total of 101 million Illumina reads were obtained from the untreated (control) and treated (bisulfite- treated) samples. Of these reads, 17% to 34% of BS-reads were successfully aligned to the converted reference genome, resulting in 10.4 to 21.7 × coverage of the 43 Mbp genome (Table 3.1, Table S32-S33).

Table 3.1 Summary of BS-Seq sequencing.

Samples	Total reads analysed	Mapped reads (percentage)	Reads unmapped (percentage)	Average length	Average coverage
Rhizines (BS-treated)	27,214,886	4,654,196 (17%)	22,560,690 (83%)	96	10.4
Rhizines (Control)	4,954,191	1,019,652 (21%)	3,934,539 (79%)	94	2.3
Thallus Body (BS-treated)	26,600,702	9,068,199 (34%)	17,532,503 (66%)	101	21.2
Thallus Body (Control)	5,639,890	2,189,990 (39%)	3,449,900 (61%)	96	4.9
Thallus Whole (BS-treated)	31,543,593	9,309,689 (30%)	22,233,883 (70%)	101	21.7
Thallus Whole (Control)	5,344,234	1,865,859 (35%)	3,478,375 (65%)	96.7	4.2

The total number of reads obtained from the three tissues rhizines, thallus body and thallus whole treated with bisulfite, and the reads from the untreated samples are indicated. Whole genome sequencing and bisulfite sequencing were carried out for the untreated and treated samples,

respectively. Mapped reads are the total number of mapped reads on the reference. Average coverage is the average coverage across all regions of the reference. The control reads were mapped using the same parameters as the BS-treated reads (see text for details).

The bisulfite conversion rate can be determined by spiking the bisulfite DNA with ~1% of lambda DNA during sample preparation. Usually, the bisulfite conversion rates are high (around 99% with the Epitect kit). In this study, the bisulfite conversion efficiency was determined by mapping the BS reads onto the *P. membranacea* mitochondrial (mt) genome (accession no: NC_016957). On average, the unconverted methylation level of Cs in the mt genome was determined to be 1.5%, 1.4% and 1.5% for CG, CHG and CHH sites, respectively. Assuming that the mtDNA is completely unmethylated, these values (false positives) were subtracted from subsequent analyses, to yield corrected/estimated numbers for CG, CHG and CHH methylation (Figure 3.4).

Methylation was found in all sequence contexts (CG, CHG, and CHH). Cytosines in CHH contexts were methylated to the greatest extent, followed by CHG and CG. A total of 12.9%, 11.2%, and 12.7% of cytosines were methylated in the rhizines, thallus body, and thallus whole tissues, respectively (Figure 3.4A and Table S34). On average, the methylation level was approximately 14%, 12.2% and 14% for CG, CHG and CHH sites, respectively. When the bulk methylation levels of rhizines and thallus body were combined and compared to thallus whole tissue they showed a similar methylation level (Figure 3.4A).

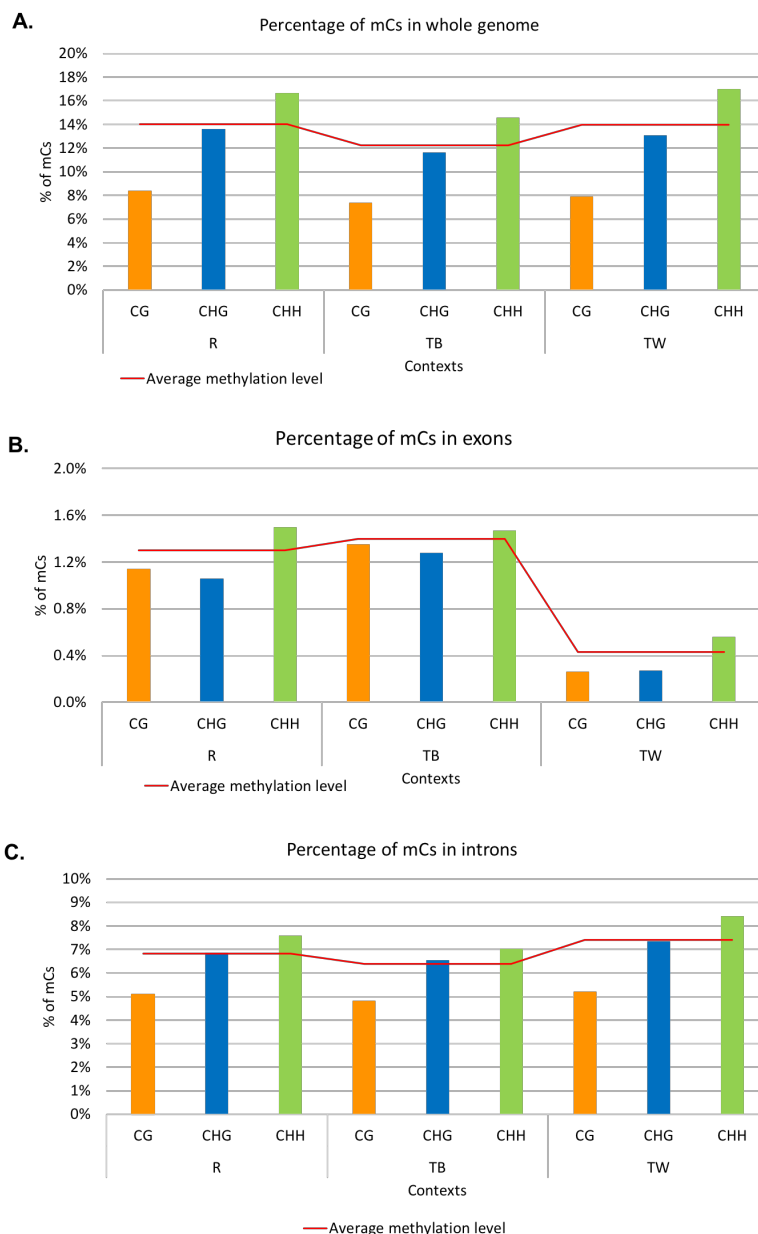


Figure 3.4 DNA cytosine methylation levels in *P. membranacea*. Percentage of methylated cytosines (mCs) identified for CG, CHH and CHG contexts in (A) whole genome (B) exons and (C) introns of the different tissues, rhizines (R), thallus body (TB) and thallus whole (TW). x- axis represents the three sequence contexts and y-axis the percentage of methylated cytosines (mCs).

Methylation has been observed in active genes (gene body methylation) of ascomycetes, e.g., in *U. reesii* (Zemach et al. 2010). The transcriptional region of the gene i.e. exons and introns together are referred to as the gene body, and gene body methylation is preferentially located in exons away from 5'UTRs and 3'UTRs and the genes are moderately expressed (Zilberman 2017). Henceforth, in this study the term exon methylation is used rather than gene body methylation. The proportion of tissue-specific methylated sites from the rhizine and thallus body tissues did not show a marked variation across sequence contexts (Figure 3.4B). On average, approximately 1.33%, 1.40% and 0.43% of cytosines were methylated in exons in R, TB, and TW, respectively (Figure 3.4B).

Tissue-specific differentially methylated regions (DMRs) in exons and introns of the rhizine and thallus body tissues (R vs. TB, TB vs. R) were identified for each sequence context (CG, CHG, and CHH) using the Fisher exact test with $p < 0.05$ significance level (see methods). Out of 16,400 genes examined DMRs in exons were observed in 103 genes (Figure 3.5, Table S35). The genes were further investigated, to see if the methylation differences coincided with gene expression differences. Negative correlation between DNA methylation and gene expression levels were observed in DMRs of exons. The list of all the genes with exon methylation and expression levels is provided in Figure 3.5 and Table S36. The methylation level and expression level for all the genes at each sequence context is shown in Figure S23. An example of exon methylation (gene 15,073) is shown in Figure 3.6.

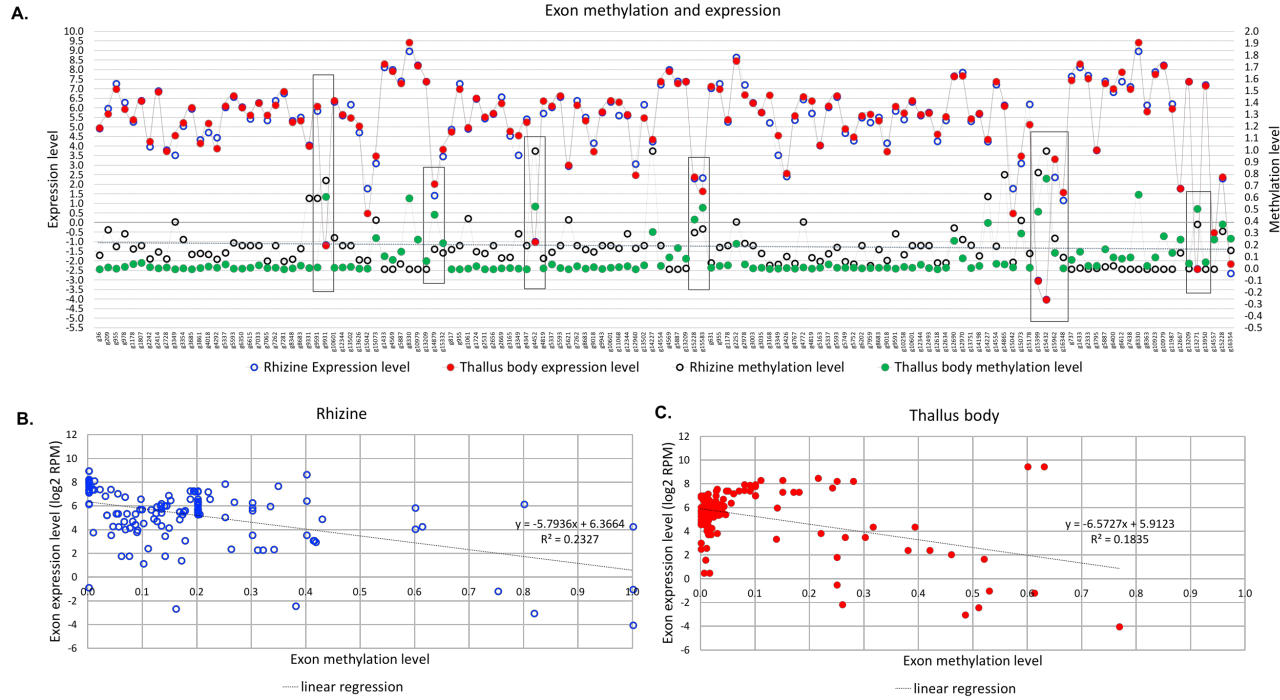


Figure 3.5 Methylation and expression levels in exons.

Line plot of methylation and expression levels in exons of rhizine and thallus body tissues. x- axis represents gene IDs, primary y-axis represents expression level ($\log_2$ RPM) and secondary y-axis methylation level. Example of genes showing a strong negative correlation are shown in black box. Scatter plot of methylation and expression levels in exons of B. rhizine and C. thallus body. x-axis and y-axis, methylation and expression levels ($\log_2$ RPM) of individual exons, respectively. Lines represent linear regression.

gypsy, and RC-Helitron were used as queries against the introns. The length of the introns ranged from >50 to <450, and the length of transposons ranged from 1,400 to 25 bases. About 28,412 introns (31.7%) had part of a TE present in them which are presumably methylated. DMRs in introns was calculated as above. Differential CG, CHG and CHH intron methylation was observed in approximately 301 scaffolds (Table S37, Figure 3.7). The list of all the scaffolds with methylation level in introns are provided in Table S38. An example of intron methylation is shown in Figure 3.8.

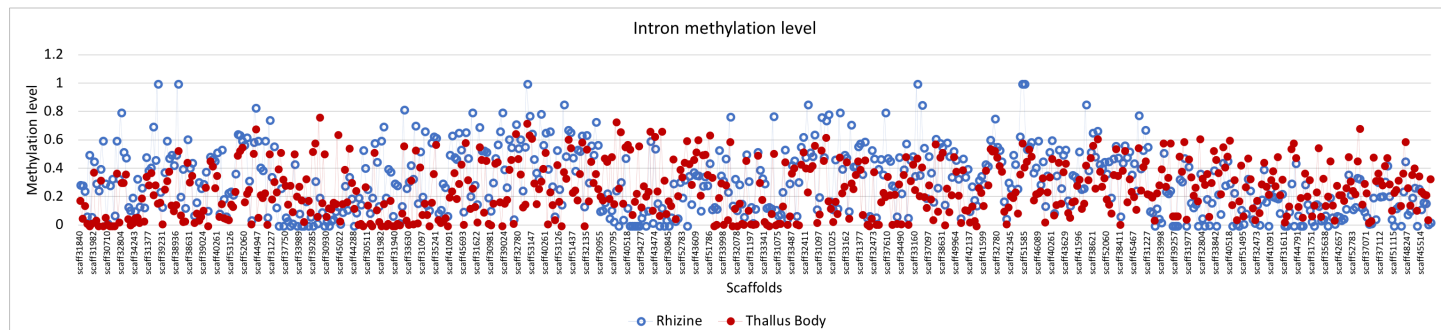
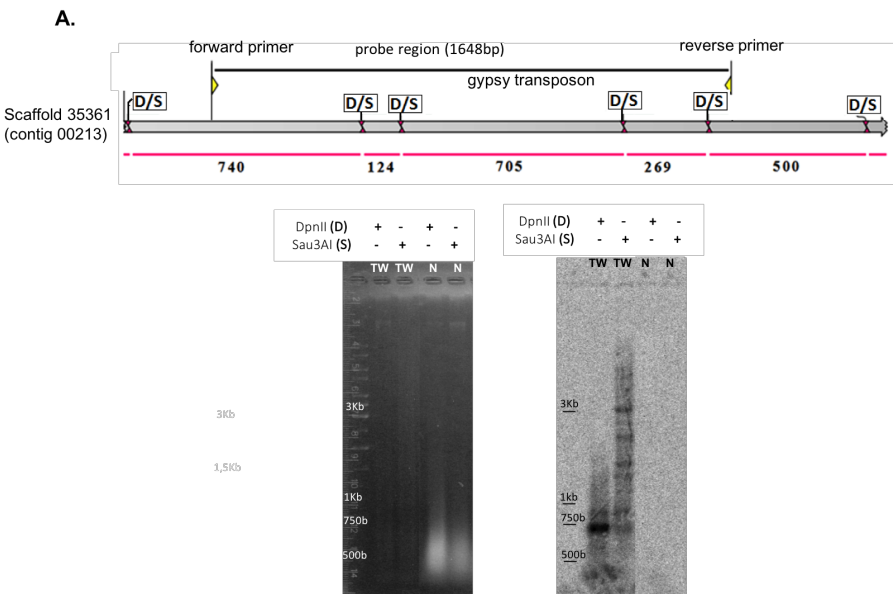


Figure 3.7 Methylation in introns.

Line plot of methylation levels in introns of rhizine and thallus body tissues. x- axis represents the scaffolds, y-axis represents the methylation level.

3.3.2 Confirmation of DNA cytosine methylation by Southern blot analysis

Southern hybridizations of *DpnII* (methylation-insensitive) and *Sau3AI* (methylation-sensitive) digested genomic DNA with radiolabeled gypsy transposon showed differences in restriction patterns between isoschizomers consistent with the presence of methylated cytosines impeding cutting in these DNA regions. DNA cleaved by *Sau3AI* showed fragments of larger sizes and greater intensity than the DNA cleaved by *DpnII*. These results indicate that the gypsy transposon in thallus DNA is methylated. When the *ded-1* gene was used as a probe, no clear differences in restriction patterns between isoschizomers were detected in apothecial or thallus DNA. The results were consistent with the presence of methylated cytosines in gypsy transposable elements (Figure 3.9A and Figure 3.10C), but not in *ded-1* gene (Figure 3.9B and Figure 3.10D).



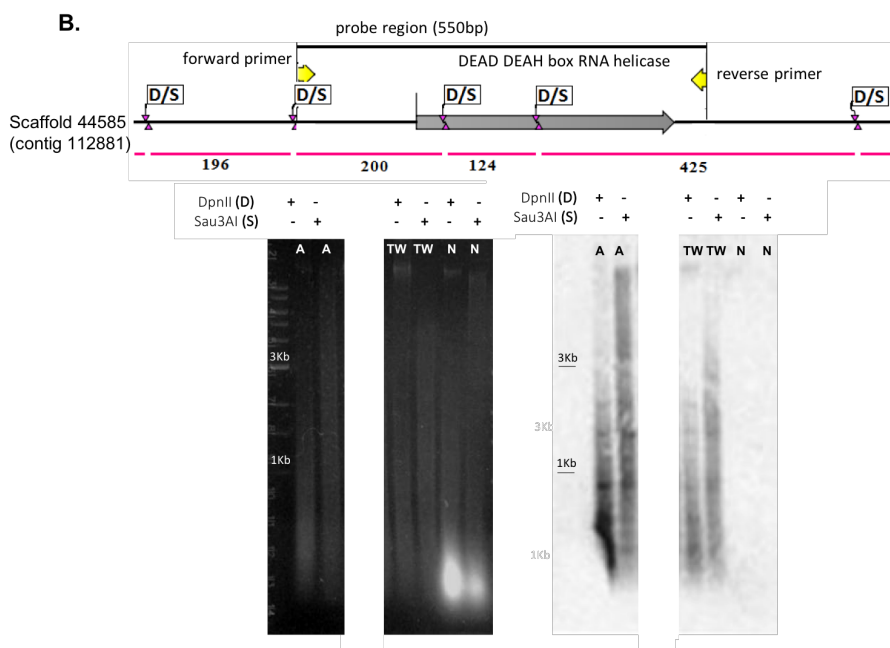


Figure 3.9 Southern blotting.

Genomic DNA from apothecia and thallus of *P. membranacea*, and *Nostoc* (negative control) were digested with isoschizomers *DpnII* (D) and *Sau3AI* (S), fractionated by gel electrophoresis followed by Southern blotting and hybridized with radiolabeled probes derived from (A) gypsy transposon and (B) *ded-1* gene. The left panel is an electrophoretic agarose gel stained with ethidium bromide, and the right panel is a Southern blot autoradiogram of the same gel.

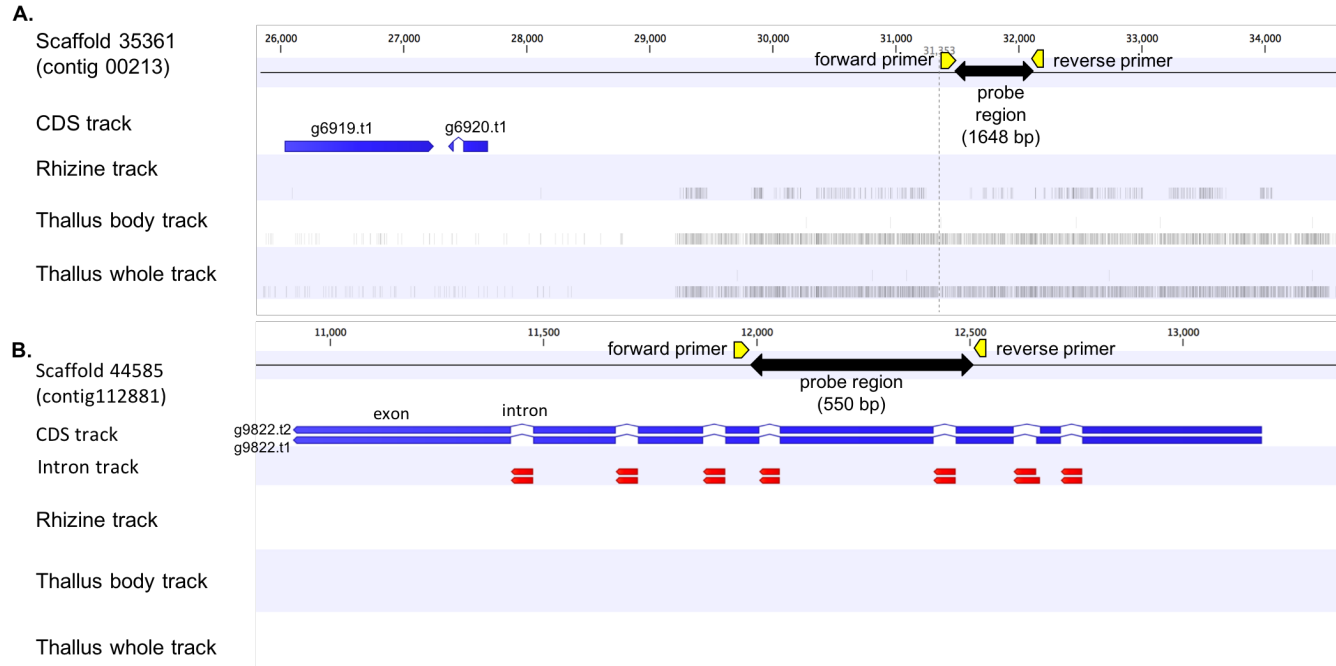


Figure 3.10 Validation of DNA methylation using Bisulfite sequencing.

Screenshot from CLC Genomics Workbench showing bisulfite-converted genomic DNA corresponding to (gypsy transposon); scaffold 35,36 (C) and *ded-1* gene; scaffold 44,585; gene 9,822 (D). Density of mC for the scaffolds, and the probe and primer regions are depicted. The isoforms are referred to as .t1 and .t2. Methylated sites are shown as individual BS-Seq tracks generated for the tissues rhizines (R), thallus whole (TW), and thallus Body (TB). Methylated cytosines are shown as vertical black lines.

3.3.3 Absence of RID and repeat–induced point mutations

RIP indices for the whole genome and repeat elements >400 bp e.g. 2205 bp from *Copia-LTR* and 1924 bp from *Gypsy-LTR* showed no evidence of RIP silencing in *P. membranacea*. RIP indices for all the transposable element families in *P. membranacea* is presented in Table 3.2 and Table S30. To date the only characterized gene involved in RIP is *rid* (RIP defective), encoding a DNA methyltransferase. Homologs of *rid* could not be found in the cyanolichen *P. membranacea* (Pmem v1.1 genome and Trinity transcriptome assemblies). Interestingly, the *rid* gene has been found in species of the subdivision Pezizomycotina, including the class Lecanoromycetes (Table S31.)

Table 3.2 RIP indices for the transposable element families in *P. membranacea* and several other ascomycetes.

	Sequences analysed	No of bases & No of sequences	Organism	TpA/ ApT *	(CpA+TpG)/ (ApC+GpT) *	References
Genome	P.mem454Allcontigs	150,581 seqs	<i>Peltigera membranacea</i>	0.85	1.23	This study
	P.mem v1.0	14,674 seqs		0.79	1.25	This study
	P.mem v1.1	3,033 seqs		0.81	1.24	This study
Transposable elements	DNA	103bp x 3	<i>Peltigera membranacea</i>	1.15	0.73	This study
	DNA-Ginger	202bp x 4		0.63	1.37	This study
	DNA-hAT	225bp x 2		0.57	1.66	This study
	DNA-hAT-Restless	126bp x 1		0.57	1.66	This study
	DNA-MuDR	194bp x 12		0.74	1.67	This study
	DNA-TcMar-Fot1	520bp x 22		0.74	1.32	This study
	DNA-TcMar-Pogo	47bp x 1		0.74	1.32	This study
	DNA-TcMar-Tc1	88bp x 2		0.67	1.38	This study
	DNA-TcMar-Mariner	75bp x 10		0.31	3.17	This study
	LINE-1	68bp x 6		0.67	2.67	This study
	LINE Penelope	62bp x 2	<i>Sclerotinia sclerotiorum</i>	0.00	0.07	This study
	LINE-Tad1	122bp x 24		0.58	1.59	This study
	LTR-Copia	2205 x 418		0.73	1.29	This study
	LTR-Gypsy	1924 x 2000		0.80	1.29	This study
	RC-Helitron	422 x 9		0.86	1.07	This study
	LTR-Copia	NA		1.35	0.27	Goldfarb et al. (2016)
	LTR-Gypsy	NA		1	0.95	
	Transposase	NA	<i>Colletotrichum cereale</i>	2	0.44	Crouch et al. (2008)
	PetTra	NA	<i>Penicillium chrysogenum</i>	1.22	0.58	Braumann et al. (2008)
	Fot1	NA	<i>Fusarium oxysporum</i>	1.12	0.75	Daboussi et al. (1992)
	Punt	NA	<i>Neurospora crassa</i>	1.32	0.56	Margolin et al. (1998)

* Standard reference values of the RIP indices are: TpA/ApT > 0.89 and (CpA + TpG)/(ApC + GpT) < 1.03. Oliver and Hane (2008)

3.3.4 DNA methylation and expression of transposable elements (TE), an inverse relationship

The DNA methylation level and expression level were examined for the transposons identified in the *P.mem* v1.1 genome. An inverse correlation between TE methylation levels and TE expression in both rhizine and thallus tissues was observed. Some of the highly expressed TEs showed a very low level of methylation (eg. g2744, Table S39). Figure 3.11 shows that the transcriptional activity and methylation levels of TEs in *P. membranacea*.

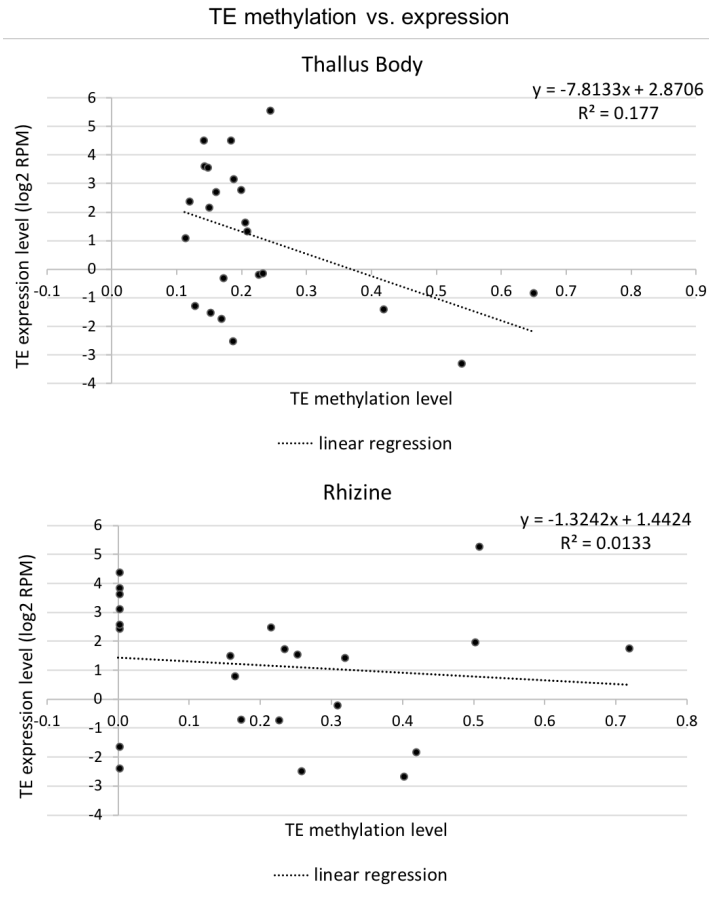


Figure 3.11 Transposon methylation and expression. x-axis and y-axis, methylation and expression levels (log₂ RPM (reads normalized per feature per million reads)) of individual TEs, respectively. Dotted lines represent linear regression.

3.3.5 RNA silencing genes

The *P. membranacea* genome contains most of the core RNAi components of quelling and MSUD, including two Dicer genes (*dcl-1* and *dcl-2*), two Argonaute genes (*qde-2* and *sms-2*) and other family members in the RNA-induced silencing (RISC) complex. The transcripts along with their expression levels are listed in Table 3.3. The presence of active core RNAi components together with a RdRP homolog as evident from the expression levels (Table 3.3) indicates a functional RNAi pathway in *P. membranacea*.

3.3.6 Identification of genes encoding DNA methyltransferases

Genes identified and thought to be involved in DNA methylation are listed in Table 3.3. DNA methyltransferase, *dim-2* (Table 3.3) was identified in the genome of *P. membranacea*. The methyltransferase was verified by aligning the predicted catalytic domain with those of other fungal methyltransferases (Figure 3.10). Based on specific motif signatures and sequence similarity to known fungal DMTs it was assigned to the DIM-2 class of DMTs. Genes involved in the DCDC and HCHC complexes were also identified in the *P. membranacea* genome. The TET proteins (TET1, TET2 and TET3) were not found in the *P. membranacea* genome.

Table 3.3 DNA methyltransferases and genes involved in RNA interference in *P. membranacea*.

Functions in <i>Neurospora crassa</i>	Gene Name	<i>N. crassa</i> accession	<i>P. membranacea</i> top hit transcript homolog (with lowest E-value)	A (log ₂ RPM)	TB (log ₂ RPM)	R (log ₂ RPM)
DNA methylation machinery						
Defective in methylation-2, Denovo CpN DMT/ maintenance CpG DMT	<i>dim-2</i>	NCU02247	g5713_scaff31383_complement (3e-140)	5.98	4.72	4.72
dim2-associated protein 1	<i>dap-1</i>	NCU04258	g3990_scaff32384_complement (2e-07)	6.55	6.06	6.04
Serine/threonine-protein phosphatase PP1	<i>ppp1</i>	NCU00043	g13450_scaff36706_complement (0.0)	6.97	6.79	6.79
SW12/SNF2-like protein. Involved in CpN methylation	<i>ddm-1</i>	NCU03875	g7959_scaff31748 (0.0)	5.04	4.35	4.47
Repeat induced point Mutations						
Putative DNA methyltransferase. Essential for RIP	<i>rid-1</i>	NCU02034	Absent [i.e. not detected.]			
DCDC complex						
Defective in methylation-7	<i>dim-7</i>	NCU04152	g12061_scaff38168_complement (5e-08)	6.27	6.02	5.89
Histone-lysine N-methyltransferase, H3 lysine-9 specific	<i>dim-5</i>	NCU04402	g15644_scaff47430_complement (2e-60)	4.90	2.59	1.87
DNA damage-binding protein 1	<i>ddb1 (dim-8)</i>	NCU06605	g6209_scaff35481 (0.0)	6.21	5.39	5.52
Cullin	<i>cul4</i>	NCU00272	g14575_scaff39159 (1e-113)	5.85	5.90	5.65
Defective in methylation-9	<i>dim-9</i>	NCU01656	g9530_scaff33771 (9e-69)	5.55	3.48	3.02
HCHC complex						
Heterochromatin protein one, essential for CpN methylation	<i>hp1</i>	NCU04017	g2032_scaff31924 (1e-33)	6.55	6.08	5.98
Chromodomain protein-2	<i>cdp-2</i>	NCU00738	g1712_scaff30495 (3.7)	5.50	5.32	5.19
Histone deacetylase	<i>had-1</i>	NCU01525	g10978_scaff39558_complement (0.0)	5.50	5.25	5.20
CDP-2 and HAD-1 associating protein	<i>chap</i>	NCU01796	g14498_scaff47729 (4.3)	5.48	5.52	5.67
Protein phosphatase PP1	<i>ppp1</i>	NCU00043	g13450_scaff36706_complement (0.0)	6.97	6.79	6.79
DNA damage checkpoint protein rad 24	<i>rftn-1</i>	NCU03300	g4314_scaff33611 (2e-79)	8.00	7.95	8.19
RNA-mediated silencing						
Quelling						
RNA-dependent RNA polymerase						
RdRP, essential for quelling	<i>qde-1</i>	NCU07534	g10029_scaff34063 (2e-98)	3.78	3.77	3.62
Sad-1	<i>sad-1</i>	NCU02178	g1467_scaff33660 (0.0)	5.19	3.49	3.79
rrp-3	<i>rrp-3</i>	NCU08435	g14034_scaff30847 (0.0)	7.02	6.16	5.98
Argonaute homologs						
Argonaute-like protein, essential for quelling	<i>qde-2</i>	NCU04730	g9628_scaff32033 (3e-165)	5.59	5.86	5.66
sms-2	<i>sms-2</i>	NCU09434	g3514_scaff30613 (0.0)	7.64	7.40	7.24
RecQ helicase homologs						
RecQ helicase, essential for quelling	<i>qde-3</i>	NCU08598	g9484_scaff51872 (0.0)	5.26	4.01	4.11
qde-3h/rqh-2	<i>recQ-2</i>	NCU03337	g7105_scaff33835 (1e-98)			
Dicer RNase homologs						
Dicer-like protein, involved in quelling	<i>sms-3 (dcl-1)</i>	NCU08270	g4270_scaff30910 (0.0)	4.99	3.74	3.72
Dicer-like protein involved in quelling	<i>dcl-2</i>	NCU06766	g1672_scaff31205 (3e-121)	5.80	5.26	4.93
Ribonuclease III-like	<i>RNase III</i>	NCU01762	g2759_scaff31488 (9e-24)	6.21	5.59	5.50
QDE-2-interacting protein, putative exonuclease protein, involved in quelling	<i>qip</i>	NCU00076	g3048_scaff31642_complement (1e-51)	5.40	5.67	5.62
MSUD						
RdRP essential for MSUD	<i>sad-1</i>	NCU02178	g1467_scaff33660 (0.0)	5.19	3.49	3.79
Essential for MSUD	<i>sad-2</i>	NCU04294	g1691_scaff31205 (1.8)	6.07	4.90	4.66
Putative eukaryotic helicase (suppressor of ascus dominance-3 protein)	<i>sad-3</i>	NCU09211	g1615_scaff33078_complement (1e-175)	5.47	4.38	3.92
Chromatin remodeling factors						
NAD-dependant histone deacetylase involved in TGS	<i>sir2</i>	NCU04737	g3608_scaff32887 (2e-171)	5.84	5.98	5.91
Histone deacetylase involved in CpG methylation	<i>had 6</i>	NCU00824	g3342_scaff33737_complement (0.0)	4.18	4.26	3.82
Chromatin remodeling complex ATPase chain ISW1	<i>crf4-1</i>	NCU03875	g7959_scaff31748 (0.0)	5.67	5.71	5.29
DNA methylation modulator						
DNA methylation modulator-1	<i>dmm-1</i>	NCU01554	g1979_scaff33213_complement (0.0)	5.17	4.70	4.85
Hypothetical protein	<i>dmm-2</i>	NCU08289	g8492_scaff38080_complement (2e-05)	4.87	5.40	5.14

Genes that display twofold expression change ($\log_2 \geq 2.0$) between samples are colour coded. Gene is coloured green if there is more than twofold expression change between samples and red colour if a gene shows less than twofold expression change.

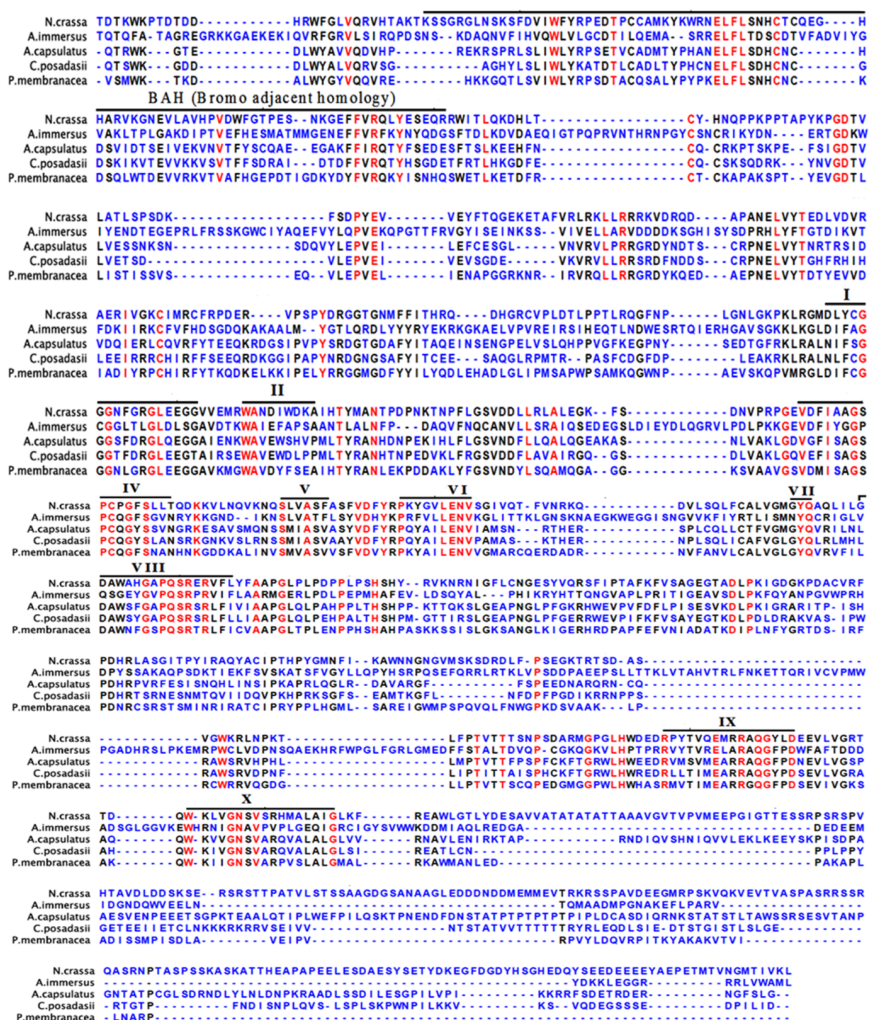


Figure 3.11 Sequence alignment of DIM-2 of *P. membranacea*. Alignment of DIM-2 from *P. membranacea*, *N. crassa* (accession no. AF348971), *Ajellomyces capsulatus* (accession no. EEH09109), *Coccidioides posadasii* (accession no. XP_003065502), and *Ascobolus immersus* MASC2 (accession No. CAB09661). Red letters indicate amino acid identity in all five sequences, black letters indicate amino acid identity in four out of five sequences, and dashes indicate gaps introduced by the CLUSTALW (Thompson et al. 1994) to maximize alignment. The ten conserved motifs defined previously based on prokaryotic MTases (Kumar et al. 1994) are indicated by Roman numerals above the aligned sequences.

3.4 Discussion

Cytosine DNA methylation is a heritable epigenetic modification important in the regulation of various processes, such as genomic imprinting, regulation of gene expression and genome defense against TE (Law and Jacobsen 2010). To date, epigenomic studies have been applied to only one lichenized fungus, *C. grayi* (Armaleo and Miao 1999; McDonald 2011). In the first study, Armaleo and Miao (1999) used restriction analysis and reported high methylation of mycobiont genes in the symbiotic tissues (poetida and soredia). Whereas in the second study, Tami McDonald (2011) used BS-Seq and reported very low methylation of mycobiont DNA in the symbiotic tissues (squamules and podetia). The above two studies produced inconsistent results, and there were certain limitations to the studies. The first study used four PKS-type fungal genes as probes. Out of the four genes, three were interpreted as being highly methylated in the symbiotic tissues whereas the fourth gene did not. Moreover, the study used the same family of genes and would have benefited by including probes such as conserved genes and repeat elements. However, considering the time when the study was undertaken this would have been difficult and time-consuming and still not as direct and robust as current methods. Also, the restriction analysis was not sensitive enough as the methylation effects in the mycobiont and photobiont of the symbiotic tissues could not be reliably differentiated. The second study (McDonald 2011) left the author surprised and perplexed as strategies were unsuccessful in confirming the high level of DNA methylation in the mycobiont present in the symbiotic state. For instance, high methylation was indicated in the symbiotic tissues using restriction analysis whereas methylation was found to be uncommon in the symbiotic mycobiont DNA using BS-Seq. Nevertheless, from the studies, it can be surmised that the high level of methylation in the symbiotic state is contributed by the photobiont (alga). In conclusion, the mycobiont in *C. grayi* has a very low or negligible level of methylation in the symbiotic state, as also seen in the aposymbiotic and cultured state.

The methylation study presented here is the first for a cyanolichen and provides new information on DNA methylation in a *Peltigera* species. Bisulfite sequencing was used to determine DNA methylomes in two distinct tissues (thallus and rhizines) of *P. membranacea* to a single base resolution. Bisulfite sequencing data in conjunction with transcriptomic data and the annotated genome of *P. membranacea* was

then used to gain a better understanding of the genes methylated in *P. membranacea*.

Cytosine DNA methylation varies significantly. For instance, a high level of DNA methylation is found in animals and plants whereas it is extremely scarce among insects and in some fungal species (e.g., *Aspergillus flavus*), and is not found in some invertebrates (Zemach and Zilberman 2010). The average methylation of cytosines in *P. membranacea* is ~13.4%, similar to that found in the plant–symbiotic black truffle (*Tuber melanosporum*). Unlike *T. melanosporum* which is extremely rich in TEs and benefits from DNA methylation to silence its repeat elements (Montanini et al. 2014), DNA methylation in *P. membranacea* is present outside REs, i.e. including introns and exons (gene body methylation). The average methylation levels of introns and exons were 6.9% and 1.1%, respectively.

The DNA methylation pattern was consistent with that found in other fungal species, and DNA methylation was observed in all sequence contexts (CG, CHG and CHH) with no preference for CG methylation. Also, the methylation level was high in non-CG contexts compared to the CG context. Comparison of rhizine and thallus tissues showed no statistically significant difference in DNA methylation. Unlike the chlorolichen *C. grayi* where methylation was uncommon in the mycobiont (McDonald 2011), the mycobiont of the cyanolichen *P. membranacea* was methylated to a high degree in both symbiotic-thallus body (~14%) and aposymbiotic- rhizine tissues (~13%).

Southern hybridizations using methylation-sensitive and insensitive restriction enzymes were carried out to determine the methylation status of specific genes in the *P. membranacea* genome. Restriction enzyme digests of genomic DNA from *P. membranacea* apothecia and thalli were probed with radiolabeled gypsy transposon and a DEAD DEAH box RNA helicase gene. The gypsy transposon appeared heavily methylated whereas the DEAD DEAH box RNA helicase gene showed no observable difference in the *DpnI* and *Sau3AI* digestion, indicating no methylation, consistent with the BS–Seq results.

Some of the highly expressed TEs listed in Table S39 were markedly less methylated in one of the two tissues (rhizine and thallus body) (e.g. g2744, g7277, g10356, and g10359). There was no significant difference in gene body methylation between the symbiotic thallus tissue

and the aposymbiotic vegetative rhizine tissue. Integration of methylation data with gene expression data provided a few insights on the role of epigenetics in *P. membranacea*. As DNA methylation often affects gene expression, differentially methylated regions (DMRs) were identified in *P. membranacea*. Approximately 12% of scaffolds showed exon and intron DMRs. The average methylation of exons for rhizine and thallus body were similar ~1.36%, whereas for thallus whole ~ 0.43%. The differences in the average methylation of exons between thallus whole and the other tissues could be attributed to the coverage of reads in the exon region. A negative correlation of exon methylation and gene expression were identified which has also been observed in other organisms (Li et al. 2008; Zilberman et al. 2007; Zhang et al. 2006; Zemach et al. 2010).

Methylation is presumably mediated by transcription factors and genome defense complexes such as the DCDC and HCHC complexes that are involved in maintaining genome integrity and regulating genome activity (Borkovich et al. 2004). Proteins of the DCDC and HCHC complexes were identified, suggesting that heterochromatin formation in *P. membranacea* is similar to what has been observed in *N. crassa* (Galagan et al. 2003a) and *S. macrospora* (Nowrousian et al. 2010).

Eukaryotes have evolved and/or retained different DMT complements. The number of predicted DMTs (Dnmt1 or Dnmt1-related) in filamentous fungi ranges from one to two, and their sequence specificity and mode of action are not yet fully understood. In *N. crassa*, cytosine methylation is established and maintained by the DIM-2 DMT, and both symmetric and asymmetric methylations are dependent on DIM-2 (Kouzminova and Selker 2001). RIP and MIP, two TE silencing mechanisms, inactivate repeated sequences (Galagan and Selker 2004; Barry et al. 1993). Both mechanisms require specific cytosine DMTs (RID/MASC1) (Freitag et al. 2002; Lee et al. 2008). RID is considered to have no DNA methyltransferase activity whereas MASC1 is a *de novo* DMT (Jeon et al. 2015). A homolog of DIM-2 was identified in *P. membranacea* whereas RID was absent. No evidence of RIP was observed in the whole genome sequence and transposons by *in silico* analyses (Table 3.2). For comparison, RIP indices from the transposons of *S. sclerotiorum* (Goldfarb et al. 2016), *Colletotrichum cereal* (Crouch et al. 2008), the PeTra transposons of *Penicillium chrysogenum* (Braumann et al. 2008), *Fot1* of *Fusarium oxysporum* (Daboussi et al. 1992) and *Punt* of *N. crassa* (Margolin et al. 1998) were used and that indicated RIP is

present. RID was identified and annotated in chlorolichens (Manoharan-Basil et al., unpublished) (Appendix: X3), suggesting that it is possible that the *rid* gene was lost in cyanolichens during their divergence from the chlorolichens.

In *N. crassa*, in addition, to RIP, two other genome defense mechanisms have been identified, MSUD and a form of RNA interference, quelling. The proteins required for both of these pathways include Argonaute, Dicer, helicases, and RdRP and are conserved in plants and fungi (Shiu et al. 2001). These genes were identified in the *P. membranacea* genome suggesting that the *P. membranacea* genome has active RNA-mediated silencing machinery and that double-stranded RNAs are efficiently processed into small RNAs capable of guiding DNA methylation in an RNA-dependent (RdDM) process (Mette et al. 2000; Teixeira et al. 2009). Despite the absence of *rid*, *P. membranacea* appears to prevent the spreading of transposons, presumably through other typical conserved eukaryotic machinery for genome maintenance, i.e. through DNA methylation-dependent regulation and uncharacterized RNAi mediated mechanisms.

Analysis of *P. membranacea* contigs (P.mem454 AllContigs) using Repeatmasker v3.3.0 (<http://www.repeatmasker.org>) with default parameters revealed that it contains 3.97% of repeats whereas P.mem v1.1 genome consists of 2.44% repeats, e.g. interspersed and simple repeats. The repeats influence the genome assembly process by resulting in mis-assemblies during the scaffolding process. Transposons that are present on the borders of sequence gaps in the draft assembly, which could not be assembled may also be missing from P.mem v1.1. The missing repeats (1.53%) from P.mem v1.1 may account for other repeats (e.g. tandem and low complexity repeats), small RNAs and TEs. In *N. crassa*, regions such as the ribosomal DNA repeats, centromeres, and telomeres that may constitute to about 2-3% of the genome may also be missing from P.mem v1.1 (Galagan et al. 2003). Hence, the percentage of repeats in the *P. membranacea* complete genome may be higher, and the total missing repeats may occupy approximately 3.5-4.5% of the whole genome sequence. Thus, it should be noted that the methylation study is not representative of all repeat regions in the genome. The issues mentioned above can be overcome by additional sequencing of the *P. membranacea* genome using the third generation long read sequencing technology such as PacBio and Oxford Nanopore or whole genome mapping technology (BioNano).

The 43 Mbp *P. membranacea* genome (P.mem v1.1) constitutes about, approximately 14.6 Mbp exons and 3 Mbp introns. From the present study, it can be surmised that approximately 5.7 Mbp of the whole genome is methylated with methylation in 0.45 Mbp of exons and 2.9 Mbp of introns. Presumably, about 1 Mbp of repeats is methylated in P.mem v1.1 genome. Furthermore, the methylation results provided here should be taken with caution, given the small sample size, uneven genome coverage and since there was no BS-Seq data available for the apothecial tissue (aposymbiotic reproductive tissue). The results should be validated in a study with a larger number of samples, with all tissues. Sequencing technology such as single-molecule real-time (SMRT) sequencing that enables direct detection of DNA methylation, without bisulfite conversion can be used to study the methylation pattern more effectively (Flusberg et al. 2010). However, the study and the data generated here will provide a good basis for future analysis.

3.5 Conclusion

This study presents the global DNA methylation profile at single nucleotide resolution for the mycobiont genome of the cyanolichen *P. membranacea*. DNA methylation occurs in the exons and introns as well as transposable elements. Cytosine methylation profiles of the two tissues (rhizine and thallus body) were similar suggesting that mycobiont methylation is common in both the symbiotic and aposymbiotic state, rather than exhibiting tissue-specific methylation. Gene expression is negatively correlated with methylated exons and TEs. This work provides new insights into the nature of DNA methylation in a cyanolichen and raises questions as to the relatively high level of methylation in non-repetitive DNA. It will also be of general interest to use it as a study model for further investigation of epigenetic modifications of DNA in lichens.

CHAPTER 4

GALECTIN-LIKE GENES: IDENTIFICATION AND EXPRESSION OF *LEC-1* AND *LEC-2* IN *PELTIGERA MEMBRANACEA*

(PAPER 1 AND PAPER 2)

PAPER 1

EXPRESSION OF LEC-1, A MYCOBIONT GENE ENCODING A GALECTIN-LIKE PROTEIN IN THE LICHEN *PELTIGERA MEMBRANACEA*

Vivian P. W. Miao, **Sheeba S. Manoharan**, Vesteynn Snaebjarnarson and Ólafur S. Andrésson. *Symbiosis* 57: 23–31. DOI 10.1007/S13199-012-0175-1

Expression of *lec-1*, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea*

Vivian P. W. Miao · Sheeba S. Manoharan ·
Vésteinn Snæbjarnarson · Ólafur S. Andrésson

Received: 14 April 2012 / Accepted: 8 June 2012
© Springer Science+Business Media B.V. 2012

Abstract The lichen *Peltigera membranacea* is an association of two symbionts: the mycobiont, a heterotrophic ascomycete, and the photobiont, a *Nostoc* sp. cyanobacterium. The *lec-1* gene, encoding a galectin-like protein, was identified in the metagenome sequence of the lichen, and analysis of the gene structure indicated that it was derived from the mycobiont. This was confirmed by transcriptomic analysis of freshly collected samples in which expression of *lec-1* was detected in apothecia and rhizines, tissues that are composed of only fungal cells. Expression of *lec-1* in portions of the lichen thallus that included both symbionts was up to two orders of magnitude greater than in tissues lacking photobionts; in contrast, expression of a housekeeping gene stayed relatively constant. These results show that *lec-1* expression is influenced by the presence of the photobiont, consistent with a model in which recognition of photobiont

ligands by lectins such as *lec-1* is a factor in the establishment or maintenance of compatible symbionts.

Keywords Lectin · Symbiosis · Mycobiont · Transcription · *lec-1* · *tub-2* · *gpd-1*

1 Introduction

Lectins are a large group of proteins of diverse structure that have the common property of binding to carbohydrates such as those found on the surface of microorganisms. Many have a well-known role in cell-cell recognition via surface molecule interactions (Sharon 2007): e.g. in legumes, lectins (phytoagglutinins) are involved in determining the compatibility of nodulation interactions with particular strains of rhizobia; in animals, other lectins (galectins) mediate immune responses (Vasta 2009). In mammals, the galectins form a well-delimited family on the basis of their shared affinity for β -galactosides and the presence of conserved residues in the carbohydrate recognition domain (CRD) (Cooper 2002; Leffler et al. 2004; Cummings and Liu 2009).

Most lectins known from fungi have been reported from basidiomycete species, often in the context of potential therapeutic or other practical applications (Singh et al. 2010a; 2010b; Lyimo et al. 2011), but studies in other taxa have provided additional perspectives (Balogh et al. 2003; Nowrousian and Cebula 2005; Silva et al. 2009). The participation of lectins as possible mediators of symbiosis in lichen-forming ascomycetes has been a long-standing interest among lichenologists (e.g. (Howe and Barrett 1970; Galun et al. 1976) and others). Since lichenized fungi retain the ability to produce and disperse apo-symbiotic ascospores (i.e. meiospores that develop in a fungal compartment from which the photosynthetic

The authors would like to dedicate this paper to the memory of Dr. Margalith Galun (1927–2012).

V. P. W. Miao (✉)
Department of Microbiology and Immunology,
University of British Columbia,
Vancouver, BC V6T 1Z3, Canada
e-mail: vmiao@telus.net

S. S. Manoharan · V. Snæbjarnarson · Ó. S. Andrésson
Department of Life and Environmental Sciences,
University of Iceland,
101 Reykjavik, Iceland

S. S. Manoharan
e-mail: ssm5@hi.is

V. Snæbjarnarson
e-mail: ves4@hi.is

Ó. S. Andrésson
e-mail: osa@hi.is

Published online: 07 July 2012

 Springer

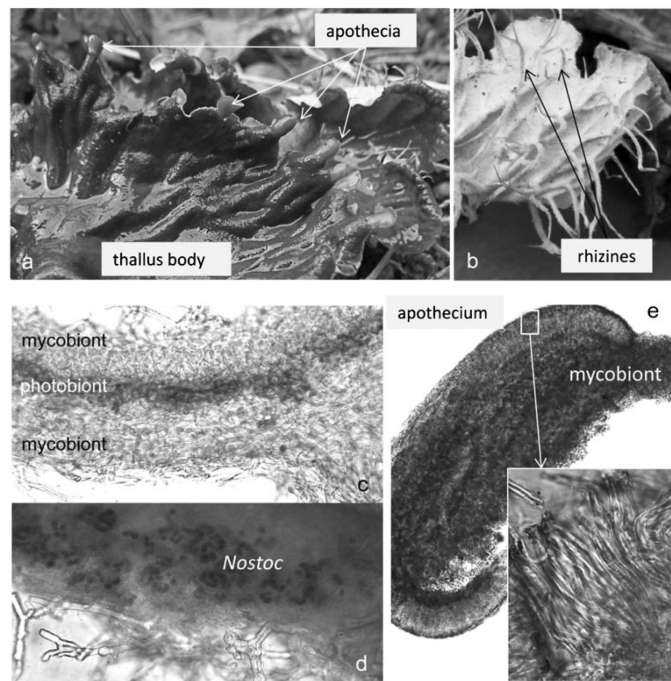
partner (the “photobiont”) is largely excluded), there must be a means for screening potential photobionts from the environment for compatibility, in order that a symbiotic relationship can be re-established *de novo*. In addition, the relationship must be maintained after initial contact, so that the stable and distinctively recognizable morphology of each lichen can be formed. Given their roles in surface recognition and specificity determination in other biological systems, it seems likely that lectins are involved at one or more stages in the interaction between symbionts in lichens.

The genus *Peltigera*, a widely distributed group of mainly terricolous foliose lichens, has provided a number of models for examining the role of lectins in lichen formation. In these lichens, the fungal partners or the “mycobionts” (referred to by the same name as the lichen) are associated with cyanobacteria (the photobiont) belonging to the genus, *Nostoc* (Fig. 1). The *Nostoc* cells are found in a layer within a thallus that is otherwise comprised of fungal cells. The latter form an upper cortex of more closely packed cells, and a medullary layer of loosely woven hyphae that give rise to

rhizines, hairlike holdfasts whose morphology and positioning are taxonomically informative. (A few species of *Peltigera* have in addition a green algal photobiont, but these lichens have a different internal arrangement). The mycobiont produces brown or blackish fruiting bodies, the apothecia, at the end of some lobes of the thallus. The apothecia contain asci and other fungal cell types and structures associated with sexual reproduction of the mycobiont, and no *Nostoc* is microscopically apparent in these tissues.

Several *Peltigera* species have been associated with the production of extracellular lectins or proteins with lectin-like properties (Lockhart et al. 1978; Petit 1982; Petit et al. 1983; Lehr et al. 1995; Diaz et al. 2009; Feoktistov et al. 2009). Some of these lectins distinguish between symbiotic and other cultured *Nostoc* sp., lending support to the idea that they play a role in the recognition process between potential partners toward establishment of the lichen. In *Peltigera membranacea*, Lehr and Galun (Lehr et al. 1995) isolated PMA (standing for *Peltigera membranacea* agglutinin), a lectin that was specifically inhibited by N-acetyl-D-galactosamine and D-galactose, and that bound to symbiotic *Nostoc*. Although

Fig. 1 Morphology of *P. membranacea*. Apothecia and upper surface of thallus body (~3.5 cm across) in a moist specimen (a) and rhizines (4–8 mm) on lower surface of the thallus body in a dried specimen (b). Cross section of thallus body (~0.4–0.5 mm thick) showing stratified arrangement of symbionts (c) and close-up showing chains of the *Nostoc* photobiont (d). Cross section of apothecium with inset showing asci and supporting cells (e)



lichen mycobionts in general are recalcitrant to culture, in some studies it was possible to confirm the mycobiont as the source of a lectin by immunoassay of germ tubes arising from ascospores (Kardish et al. 1991) or of cultured mycobionts (Bubrick et al. 1981); in *P. membranacea*, however, failure of ascospores to germinate prevented confirmation of the mycobiont as the producer of PMA.

Here we extend the investigation of lectins in *P. membranacea* as part of a whole genome sequencing project (“Pmb-WGS”) and show by transcriptome analysis of field samples that a mycobiont gene encoding a galectin-like protein is differentially expressed in symbiotic and aposymbiotic tissues.

2 Materials and methods

2.1 Bioinformatic identification of genes

The *lec-1* gene was identified in the on-going *P. membranacea* whole genome sequence (Pmb-WGS) project (unpublished, Xavier et al. 2012) using the sequences of PMA oligopeptides as queries. The database is comprised of *P. membranacea* apothecial and thallus metagenomic DNAs that were processed for sequencing at commercial facilities via the Roche 454 and the Illumina/Solexa methodologies, resulting in 1.76 Gb of 454 data and 1.4 Gb of Illumina data that were assembled using automated procedures. The database was queried using reverse translations of the PMA peptides I {(K/R)AVYSNPVVIQVFDNGTAAGPPT}, II {(K/R)SNYHNIYV(F/I)DV(I/F)} and III {(K/R)IYTFDKR} (Lehr et al. 1995) and the CLC Genomics Workbench (CLC bio, Denmark) suite of tools to identify potential matches; the relevant regions of these contigs were then examined in detail for annotation. Mycobiont “house-keeping” genes encoding the β subunit of fungal tubulin and the enzyme glyceraldehyde-3-phosphate dehydrogenase, were found by querying the Pmb-WGS with related sequences from *Aspergillus* spp. (accession P22012, P20445 respectively). Candidate genes were annotated by comparison with known genes and inspection for conserved dinucleotides (GT/AG) at prospective splice sites associated with appropriate frame shifts, and deduced features were confirmed by RNA-seq (below). The genetic nomenclature convention of *Neurospora crassa* was used in the naming of genes. DNA sequences of *lec-1*, *tub-2* and *gpd-1* were deposited in GenBank as accessions JQ899138, JQ837249 and JQ837250, respectively.

2.2 Analysis of gene expression in vivo

Lichens were collected directly into RNALater (Ambion) from one site (Keldur, Reykjavik, Iceland, 41° 64.7'N, 21°

46.6'W) in the fall 2009 and 2010, and processed within hours of collection by flash freezing and grinding under liquid nitrogen using a mortar and pestle. Total RNA was extracted from apothecia (about 50) or from ~100 mg ground wet weight of lichen using the ToTALLY RNA Kit (Ambion) according to manufacturer's instructions. RNA quality was evaluated using the Agilent 2100 Bioanalyzer RNA Nano Chip. In total, five samples were prepared: (1) basal and (2) distal halves of vegetative lobes (“thallus bases and tips”) from one lot of thalli; (3) apothecia and (4) the rest of the lobes (from which the apothecia originated) from thalli collected on a second date; and (5) rhizines from thalli obtained on a third collection date.

The 8–9 μ g total RNA obtained from each sample was prepared using an Illumina standard kit for mRNA-seq according to the manufacturer's protocol. Each library was run on a Genome Analyzer (GAIIx) lane using 35 base sequencing for apothecia and thallus, 75 base sequencing for rhizines and 100 base sequencing for thallus bases and tips. The reads were aligned to the Pmb-WGS dataset using the short read aligner Bowtie v 0.12.7 (Langmead et al. 2009) incorporated in TopHat v1.3.0 (Trapnell et al. 2009). Default parameters were used for mapping, except for the minimum (30) and maximum (300) insert size. Cufflinks v1.3.0 (Trapnell et al. 2010; Roberts et al. 2011a; Roberts et al. 2011b) was used to assemble the aligned reads into transcripts and provide an approximation of their abundance in comparison to a DNA genome reference. Trinity (Grabherr et al. 2011), a different method that is based on de novo reconstruction, was employed for confirmation of transcripts. Low quality bases and adaptor sequences were trimmed prior to assembly in all cases. The annotated transcript (spliced exonic regions) plus 10 terminal nucleotides upstream and downstream of each gene was used as a reference base to determine expression levels using the unit, FPKM (fragments per kilobase of transcript per million fragments mapped). This unit introduces a means to normalize data with respect to the total size of the sample provided for sequencing (so large and small samples can be compared) as well as the size of the genes of interest (so long and short genes can be compared).

2.2.1 Characterization of *lec-1* in other populations

P. membranacea was collected from the same site in Iceland as used for the Pmb-WGS, Keldur, (sample XBB004) and at three sites in British Columbia, Canada: Eagle Ridge, Vancouver 49° 22'N, 123° 16'W (sample VM5082), and Sooke Basin 48° 23'N, 123° 28'W and the Saanich Peninsula 48° 41'N, 123° 25'W on Vancouver Island (VM5072, VM5084, respectively). Total lichen DNA was extracted using published methods in which samples ground in liquid nitrogen are treated with detergent, chloroform,

RNAse and isopropanol or ethanol precipitation (Sinnemann et al. 2000). The *lec-1* gene was amplified with primers O-720 (CAAGTCTGATTATCTCATCC) and O-723 (CGAAACAGAGCCAGCATTTAGC) in a 3-step PCR program: 96 °C 10 s, 49 °C 10 s and 72 °C for 30 s; in some cases a second round of PCR was conducted using an internal primer, O-722 (TCCTCTGTTCCGCGCACGAC). Products were sequenced directly using the PCR primers.

3 Results

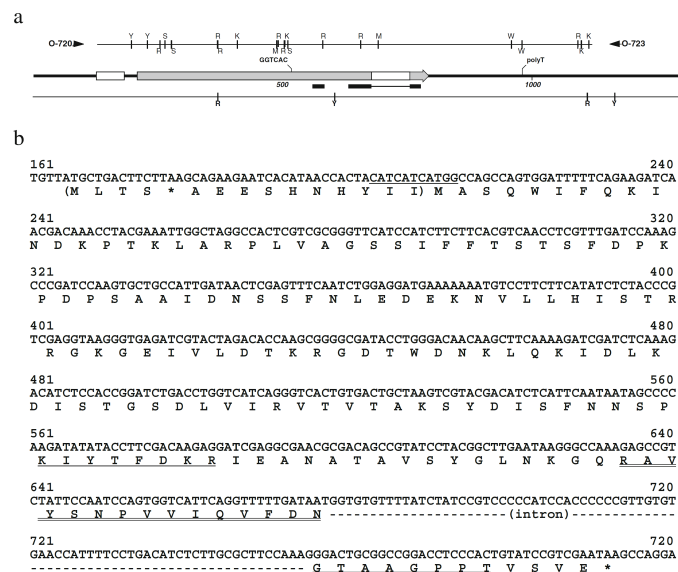
3.1 Identification and characterization of *lec-1*

Three oligopeptide sequences from *P. membranacea* lectin PMA (peptides I, II and III) (Lehr et al. 1995) were reverse translated to query the partially annotated *P. membranacea* whole genome sequencing dataset, Pmb-WGS (unpublished). A 1.3 kb region including an open reading frame with sequences corresponding to peptides I and III was found (Fig. 2). Comparison with sequences in GenBank suggested this region encoded a protein with ~30 % amino acid identity to a number of putative or proven fungal galectins (e.g. accessions XP_001877242, AAB06178). A

single interruption and frame shift of the region of homology coincided nearly exactly with a 77-nucleotide (nt) insertion in the region corresponding to the sequence for peptide I; interpretation of this intervening sequence as an intron was consistent with the presence of typical fungal intron border junctions (GT and AG). This model was supported by transcriptome data that revealed a 162 nt 5' UTR interrupted by an intron (56 nt), a 584 nt region with two coding sequences (469 nt, 38 nt) and one intron (77 nt), and a longer (400 nt) 3'UTR. *In silico* splicing of the coding sequences restored the reading frame and predicted a 168-residue protein of 18.5 kDa. A potential Kozak consensus sequence was found around the ATG (Kozak 2002). The gene identified was designated *lec-1* and its predicted product, LEC-1. Two single nucleotide polymorphisms (SNPs) in the first coding region were evident in the Pmb-WGS data, but both were silent.

LEC-1 had a match to fungal galectins by BLAST analysis and was most comparable (30 % identity) to a *Laccaria bicolor* protein (XP_001877242), with 76 % coverage and an E-value of 2e-09. Alignment to related fungal proteins showed that some of the conserved ligand-coordinating residues in the galectin carbohydrate recognition domains (CRD) were present in LEC-1, but that others have been replaced by

Fig. 2 Map of the *lec-1* gene (a) and translation product (b). Features in (a): CDS (grey), introns (white), 5' and 3' UTRs (thick lines), coordinates (grey numerals). Positions of PMA peptides I and III are shown as short line segments under the CDS (peptide I is interrupted by an intron). Positions of allelic variation: "GGTCAC" insert and poly "T" are shown on the main features; SNPs (ticks labeled with standard I letter nucleotide code denoting ambiguity) in Pmb-WGS are marked on a line below the main features; SNPs in amplicons from BC samples obtained with primers O-720, O-723 (arrowheads) are on a line above. A putative Kozak sequence around the start codon is underlined in (b), and a N-terminal tract with similarity to PMA peptide II is enclosed parenthetically. Residues for PMA peptides I and III are underlined (double and single, respectively)



		H ⁴⁴	R ⁴⁹	V ⁵⁵ N ⁶¹	W ⁶⁸	E ⁷¹ R ⁷³	
LEC-1	56	NVLLHISTR	RKG	IVL	TKR	GDTWDN	KDID 88
ACG	58	AYLLHIAF	RQENV	VFNSRQ	PNAPWL	VEQRVS	90
AAL	55	AYLLHIAF	RQENV	VFNSRQ	PDGPWL	VEQRVS	87
CGL1	47	NILLHIS	IRPGENA	IVFNSR	TGGAWG	PEERIP	79
CGL2	47	NILLHIS	IRPGENA	IVFNSR	LKNGAWG	PEERIP	79
CGL3	56	DVLLHIG	IRRRENA	VFNSIP	YCGE	GC-PEERIP	88
LAG	54	DILL	ISIRRAENA	IVMNSL	PASGNWG	TEESVP	86

Fig. 3 Carbohydrate binding domains of LEC-1 and fungal galectins. Galectin signature residues are identified at the top (human galectin-1 numbering (Hirabayashi and Kasai 1988)), inverse letters identify substitutions at these positions. ACG (*Agroclybe cylindrica*, Yagi et al. 2001), AAL (*Agroclybe aegerita* AY264782), CGL1, CGL2, CLG3 (*Coprinopsis cinerea* XP_001836010, XP_001836012, XP_001829544, respectively), LAG (*Laccaria amethystina* BAJ72707). Native amino acid position numbers are given the ends of each peptide

potentially significant substitutions (Fig. 3). Substitution for R73 (human galectin-1 numbering (Hirabayashi and Kasai 1988)) has been observed in the *Laccaria amethystina* galectin LAG, but changes at N61 and E71 have not been reported in fungi.

3.2 Identification and characterization of *gpd-1* and *tub-2*

Housekeeping genes such as those encoding β -tubulin or the enzyme glyceraldehyde-3-phosphate dehydrogenase are often used as a reference base to accompany investigations in gene expression (Bütehorn et al. 1999; Delp et al. 2003; Joneson et al. 2011). To this end, the complete *P. membranacea* homologs of these genes were identified in the Pmb-WGS. The *P. membranacea gpd-1* gene includes a pre-mRNA transcript of 1,703 nt with 7 exons and 6 introns. The predicted 339 amino acid protein, GPD-1, has >80 % sequence identity (>90 % similarity) with fungal glyceraldehyde-3-phosphate dehydrogenases (e.g. GenBank accession P54118 or P29497). The *P. membranacea tub-2* gene has a 2,235 nt transcript with 9 exons and 8 introns in the coding region. The predicted 448 residue protein, TUB-2, has >90 % sequence identity (>96 % similarity) with fungal beta-tubulins (e.g. GenBank accession P05220). In contrast to *lec-1*, neither *gpd-1* nor *tub-2* exhibited evidence of sequence variation in the Pmb-WGS.

3.3 Expression of *lec-1*

Transcripts of all three genes, *lec-1*, *tub-2* and *gpd-1* were detected in freshly collected lichens, but the levels were not the same among the thallus (here referring to the main body of the lichen, with both mycobiont and photobiont cells, and the rhizines), apothecia and rhizine samples. The ratio of *gpd-1* to *tub-2* expression, which provides a basis of comparison between samples, stayed between 1 and 2 in all three thallus-derived data sets, showing a level of reproducibility in the method (Table 1). The expression of *lec-1* was

approximately twice that of *tub-2*, and up to two-fold that of *gpd-1*, in datasets 1 and 2, representing basal and distal (“bases” and “tips”, respectively) portions of the thalli. This was also observed in dataset 3, representing thalli (bases together with tips) from which the apothecia had been removed. The trend was dramatically changed when apothecia from those same thalli were processed and analysed separately (dataset 4): while the *gpd-1* to *tub-2* ratio increased slightly, the *lec-1* ratio to both genes decreased by two orders of magnitude. Since the thallus consists of symbiotic tissue with both mycobiont and photobiont, whereas the apothecia contain only the mycobiont, elevated expression of *lec-1* in thalli may be related to a role in symbiosis. Consistent with this hypothesis, expression of *lec-1* in rhizines, another exclusively fungal tissue (dataset 5), was also reduced albeit to a lesser extent; for comparison, the *gpd-1* to *tub-2* ratio remained between 1 and 2 as before.

3.4 Allelic variation of *lec-1*

The occurrence of *lec-1* in other populations of *P. membranacea* was investigated by PCR in specimens collected from British Columbia (BC), and in another Icelandic collection, XBB004, from the same location as the source material for Pmb-WGS (Fig. 4). The sequence of the XBB004 *lec-1* amplicon was identical to that obtained from Pmb-WGS, with the exception of an additional T in a homopolymeric region in the 3'UTR (i.e. Pmb-WGS C₄A₂T₁₁, compared to XBB004 C₄A₂T₁₂). The BC amplicons were highly similar to each other and to that of the Icelandic samples (>98 % DNA identity), but were also slightly different in the polyT region: (VM5082 C₄A₁T₁₃, VM5072 C₄A₁T₁₄, VM5084 C₄A₁T₁₅). As a group, the *lec-1* alleles from the three BC samples exhibited a total of 21 SNPs, 20 of which were not observed in the Pmb-WGS (Fig. 2). Collectively, these were responsible for ten substitutions in the deduced protein, including four that were common to all 3 samples. The deduced protein from *P. membranacea* VM5072 *lec-1* also had an internal repeat that generated a two amino acid insertion. Interestingly, none of the variants detected affected the key residues in the CRD.

4 Discussion

In this study, we describe the *lec-1* gene from the mycobiont of the lichen *P. membranacea*. While both fungal and cyanobacterial sequences are included in the Pmb-WGS, the typical eukaryotic sequence features, including the presence of an intron and fungal splicing signals, rule out the photobiont as the source of *lec-1*. To date, no genes similar to *lec-1* have been found either in the *P. membranacea* *Nostoc* genome sequences, or in publicly available *Nostoc* or

Table 1 *lec-1*, *tub-2* and *gpd-1* transcription (FPKM^a) in *P. membranacea*

Dataset	Tissue	Expression			Relative Expression Ratio		
		(kFPKM)*			<i>lec-1/</i>	<i>lec-1/</i>	<i>gpd-1/</i>
		<i>tub-2</i>	<i>gpd-1</i>	<i>lec-1</i>	<i>tub-2</i>	<i>gpd-1</i>	<i>tub-2</i>
1	thallus base	168	252	313	1.86	1.24	1.50
2	thallus tip	152	220	386	2.54	1.75	1.45
3	thallus	93	180	455	4.90	2.53	1.94
4	apothecia	158	366	2	0.01	<0.01	2.32
5	rhizines	213	336	93	0.44	0.28	1.57

^a thousand Fragments Per Kb of transcript per Million fragments mapped

Anabaena genomes. Subsequent detection of *lec-1* polyA mRNA in fungal tissues also supports the conclusion that this gene is derived from the mycobiont genome.

The *lec-1* gene was initially identified by matching the partially annotated whole genome sequence of *P. membranacea* to two of three reverse-translated peptides reported from PMA, an N-acetyl-D-galactosamine reactive lectin characterized from *P. membranacea* (Lehr et al. 1995). The absence of a clear match for PMA peptide II in the Pmb-WGS could be due to genetic differences in the lichen populations used for Pmb-WGS and for the PMA studies; detection of a variety of *lec-1* alleles with non-synonymous substitutions and insertions in the BC populations are consistent with this possibility. Many fungal studies have reported isolectins and isoforms of galectins from one species (e.g. Yagi et al. 2001; Lyimo et al. 2011). In some cases, they may be due to point mutations, e.g. *C. cinerea* galectin CGL2 (e.g. AAB06178, XP_001836012), but in other cases, there may be more than one gene (Iijima, 2002, Trigueros, 2003]. In the Pmb-WGS, *lec-1* was the only possible match for PMA and all reads assembled to one contig; therefore, barring large duplications of the genome, *lec-1* appears to be in single copy. Two observations suggest that PMA might be an N-terminally extended version of LEC-1: (1) there is a resemblance between the middle residues of the unmatched PMA peptide II (K/R•S•N•Y•H•N•I•Y•V•I/F•D•V•I/F) and a theoretical, immediate upstream extension of LEC-1 (H•N•H•Y•I•I), and (2) there is an ~1.5 kDa difference between the observed molecular weight of ~20.1 kDa for

PMA (see (Lehr et al. 1995), Fig. 1) and that predicted for LEC-1 (18.5 kDa) that is similar to the estimated mass range of peptide II (1.56 kDa to 1.65 kDa, owing to ambiguity at three positions). A putative extended form might arise as an allelic variant, but it should be noted that in *Arthrobotry oligospora*, only one gene for lectin AOL was detected by Southern hybridization, and a post translational modification was proposed to account for a detected isoform (Rosen et al. 1996). These observations are circumstantial until further biochemical characterization of LEC-1 allows a full comparison to PMA, but nonetheless, the PMA-containing population may prove worthwhile to investigate for novel *lec-1* alleles, particularly with respect to examination of the CRD. Although the best matches for LEC-1 among sequenced proteins include validated fungal galectins such as CGL1 and CGL2 (Cooper et al. 1997; Walser et al. 2004; 2005), neither PMA nor LEC-1 can be placed in the galectin class at this time: the ligand inhibition phenotype of PMA is that of a galectin, but the criteria for CRD conservation have not been tested, while despite the sequence similarity of LEC-1 to galectins, it falls short in the conserved CRD motifs and its β-galactoside binding profile has not been examined. Thus it is recommended that they be designated more generally as galectin-like proteins (Leffler et al. 2004). The findings from this study suggest that in retrospect, PMA was also of mycobiont origin, and review of its reported properties (Lehr et al. 1995) might benefit from this perspective. Interestingly, *lec-1* may be the first ascomycete galectin-like protein gene

Fig. 4 *P. membranacea lec-1* allelic variation. Icelandic sample XBB004 is used as the reference; divergent residues in BC samples are indicated by *inverse color*. Residues belonging to the carbohydrate binding domain are in bold, and amino acid positions are shown above



reported: although a number of ascomycete hits were recognized in GenBank by BLASTP analysis, in each case the region of similarity was a domain within a much larger protein, rather than a small independent protein comparable to the prototype galectins reported in *C. cinerea* (Cooper et al. 1997).

Expression of *lec-1* was examined using a meta-transcriptomic approach in different tissues and samples. This permits the identification of individual or suites of co-expressed genes whose temporal or spatial/tissue specific activity profile may be indicative of their biological roles: e.g. genes involved in symbiosis might be differentially expressed in symbiotic and apo-symbiotic portions of the thallus, while housekeeping gene expression might be similar in most tissues. FPKM values (Table 1) can be compared within a dataset with reference to a specific gene (e.g. *tub-2*), and then these ratios can be compared between datasets. In this way, differences due to the composition of source material and to technical handling of samples can be minimized, and a broader range of comparisons can be justified, between different tissues and even different taxa. The *P. membranacea tub-2* and *gpd-1* genes, interpreted as mycobiont genes based on their features and expression levels in fungal-only tissues, are typical of the reference genes used for comparative purposes (Büthorn et al. 1999; Delp et al. 2003; Joneson et al. 2011). Inclusion of both as comparators increases confidence in the event of unanticipated influences on one of the references (Shinohara et al. 1998). In this study, *lec-1* expression patterns were clearly distinct from those of *gpd-1* and *tub-2*, suggesting different roles.

A variety of functions have been proposed or demonstrated for different lectins. Where they have been examined in a tissue-specific context, expression of basidiomycete lectins is usually up-regulated with fruiting body development (Oguri et al. 1996; Boulianne et al. 2000; Sun et al. 2003). In the ascomycete *Sordaria macrospora*, expression of the lectin gene *tap-1* was also up-regulated during sexual development, albeit not being essential for fruiting (Nowrousian and Cebula 2005).

In *P. membranacea*, the most notable tissue-specific correlation with *lec-1* expression is whether the tissue includes the photobiont. One possibility consistent with the observed differential expression in symbiotic versus apo-symbiotic tissue is that lectins such as LEC-1 may be a factor in recognition of ligands produced by the *Nostoc* photobiont, as proposed previously (Lehr et al. 1995); however, it is yet unclear if *lec-1* expression might be induced or maintained (or both), in response to putative photobiont signals. Any model for the function of LEC-1 must in addition accommodate its high level of expression in thalli, as products of highly expressed genes are often structural rather than regulatory. Given the tendency for galectins and galectin-like

proteins to function as dimers, an alternate or additional hypothesis is that LEC-1 could participate in organizing the extracellular matrix and maintaining the stratified arrangement with photobionts within the medulla. It is interesting that there were no closely related genes from other whole genome databases (at present not including any lichen symbionts), suggesting that whatever the specific function(s), it does not have a near counterpart in currently sequenced ascomycetes. A role in recognition between trees and their fungal symbionts has previously been proposed for mushroom lectins (Guillot et al. 1991; Giollant et al. 1993), and *lecA*, a lectin gene from *Paxillus involutus*, a mycorrhizal basidiomycete, has been recognized as a symbiosis-regulated gene by microarray experiments (Le Quere et al. 2006). It may ultimately prove significant that the most similar match to LEC-1 was a putative galectin from *L. bicolor*, another symbiotic basidiomycete that forms ectomycorrhizal associations with a range of temperate trees species, but many more fungal lectins will need to be sequenced and studied.

There have been no systematic surveys of allelic variation in fungal lectins even though isolectins have long been known. The SNPs and insertion in the BC population may be useful to investigate further for their influence on the specificity of ligand-binding and relative affinities, and related to this, their potential contribution to interactions with photobionts (e.g. selectivity). Experimental manipulations showed that the W68R substitution in the CRD of *C. cinerea* lectin CGL3 is critical in determining chito oligosaccharide versus lactose binding (Walti et al. 2008), while X-ray crystallographic studies of the *Agrocybe cylindricea* lectin ACG revealed that non-CRD residues are necessary for specifying sialic acid binding (Ban et al. 2005). The divergence among the *lec-1* alleles may also prove useful for phylogeography and evolutionary studies at some taxonomic levels.

A search of the Pmb-WGS using *lec-1* as a query sequence has identified additional candidate genes encoding lectins or lectin-like enzymes in *P. membranacea* that will be the subject of other investigations. Enzymes (arginase) with lectin-like properties have also been isolated from *Peltigera canina* (Diaz et al. 2009) and other cyanolichens (Sacristan et al. 2007; Vivas et al. 2010), as well as from chlorolichens (Legaz et al. 2004), and proposed to act as a recognition factor for compatible photobionts. These observations, together with the results of the present study, strongly indicate that further exploration of lectin-encoding genes will lead to a greater understanding of the factors involved in establishment and maintenance of lichen symbiosis.

Acknowledgements We thank T. Taylor for help with collection of specimens in BC, B.B.Xavier, Z.O. Jónsson and H. Jónsson for

assisting with bioinformatics, and deCode Genetics and D.N. Magnúsdóttir, G.P. Örlýgssdóttir, S. Snorraddóttir and Ó. Þ. Magnússon for technical support.

References

- Balogh J, Tunlid A, Rosen S (2003) Deletion of a lectin gene does not affect the phenotype of the nematode-trapping fungus *Arthrotrix oligospora*. *Fungal Genet Biol* 39:128–135
- Ban M, Yoon HJ, Demirkan E, Utsumi S, Mikami B, Yagi F (2005) Structural basis of a fungal galectin from *Agrocybe cylindracea* for recognizing sialoconjugate. *J Mol Biol* 351:695–706
- Boulianne RP, Liu Y, Aebi M, Lu BC, Kues U (2000) Fruiting body development in *Coprinus cinereus*: regulated expression of two galectins secreted by a non-classical pathway. *Microbiology* 146 (Pt 8):1841–1853
- Bubrick P, Galun M, Frensdorff A (1981) Proteins from the lichen *Xanthoria parietina* which bind to phycobiont cell walls. Localization in the intact lichen and cultured mycobiont. *Protoplasma*: 207–211
- Büthorn B, Gianinazzi-Pearson V, Franken P (1999) Quantification of β -tubulin RNA expression during asymbiotic and symbiotic development of the arbuscular mycorrhizal fungus *Glomus mosseae*. *Mycol Res* 103:360–364
- Cooper DN (2002) Galectinomics: finding themes in complexity. *Biochim Biophys Acta* 1572:209–231
- Cooper DN, Boulianne RP, Charlton S, Farrell EM, Sucher A, Lu BC (1997) Fungal galectins, sequence and specificity of two isolectins from *Coprinus cinereus*. *J Biol Chem* 272:1514–1521
- Cummings RD, Liu FT (2009) Galectins. In: Varki A, Cummings RD, Esko JD (eds) *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Delp G, Timonen S, Rosewarne GM, Barker SJ, Smith S (2003) Differential expression of *Glomus intraradices* genes in external mycelium and mycorrhizal roots of tomato and barley. *Mycol Res* 107:1083–1093
- Diaz EM, Sacristan M, Legaz ME, Vicente C (2009) Isolation and characterization of a cyanobacterium-binding protein and its cell wall receptor in the lichen *Peltigera canina*. *Plant Signal Behav* 4:598–603
- Feoktistov AS, Kitashov AV, Lobakova ES (2009) The characterization of lectins from the tripartite lichen *Peltigera aphthosa* (L.) Willd. *Mosc Univ Biol Sci Bull* 64:23–27
- Galun M, Braun A, Frensdorff A, Galun E (1976) Hyphal walls of isolated lichen fungi: autoradiographic localization of precursor incorporation and binding of fluorescein-conjugated lectins. *Arch Microbiol* 108:9–16
- Giollant M, Guillot J, Damez M, Dusser M, Didier P, Didier E (1993) Characterization of a lectin from *Lactarius deterrimus* (Research on the possible involvement of the fungal lectin in recognition between mushroom and spruce during the early stages of mycorrhizae formation). *Plant Physiol* 101:513–522
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Guillot J, Giollant M, Damez M, Dusser M (1991) Isolation and characterization of a lectin from the mushroom, *Lactarius deliciosus*. *J Biochem* 109:840–845
- Hirabayashi J, Kasai K (1988) Complete amino acid sequence of a beta-galactoside-binding lectin from human placenta. *J Biochem* 104:1–4
- Howe ML, Barrett JT (1970) Studies on a hemagglutinin from the lichen *Parmelia michauxiana*. *Biochim Biophys Acta* 215:97–104
- Joneson S, Armaleo D, Lutzoni F (2011) Fungal and algal gene expression in early developmental stages of lichen-symbiosis. *Mycologia* 103:291–306
- Kardish N, Silberstein L, Fleminger G, Galun M (1991) Lectin from the lichen *Nephroma laevigatum* Ach. *Symbiosis* 11:47–62
- Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299:1–34
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25
- Le Quere A, Eriksen KA, Rajashekar B, Schutzendubel A, Canback B, Johansson T, Tunlid A (2006) Screening for rapidly evolving genes in the ectomycorrhizal fungus *Paxillus involutus* using cDNA microarrays. *Mol Ecol* 15:535–550
- Leffler H, Carlsson S, Hedlund M, Qian Y, Poirier F (2004) Introduction to galectins. *Glycoconj J* 19:433–440
- Legaz ME, Fontaniella B, Millanes AM, Vicente C (2004) Secreted arginases from phylogenetically farrelated lichen species act as cross-recognition factors for two different algal cells. *Eur J Cell Biol* 83:435–446
- Lehr H, Flemminger G, Galun M (1995) Lectin from the lichen *Peltigera membranacea* (Ach.) Nyl.: characterization and function. *Symbiosis* 18:1–13
- Lockhart CM, Rowell P, Stewart WDP (1978) Phytohaemagglutinins from the nitrogen-fixing lichens *Peltigera canina* and *P. polydactyla*. *FEMS Microbiol Lett* 3:127–130
- Lyimo B, Yagi F, Minami Y (2011) Primary structure and specificity of a new member of galectin family from the Amethyst deceiver mushroom *Laccaria amethystina*. *Biosci Biotechnol Biochem* 75:62–69
- Nowrouzian M, Cebula P (2005) The gene for a lectin-like protein is transcriptionally activated during sexual development, but is not essential for fruiting body formation in the filamentous fungus *Sordaria macrospora*. *BMC Microbiol* 5:64
- Oguri S, Ando A, Nagata Y (1996) A novel developmental stage-specific lectin of the basidiomycete *Pleurotus cornucopiae*. *J Bacteriol* 178:5692–5698
- Petit P (1982) Phytolectins from the nitrogen-fixing lichen *Peltigera horizontalis*: the binding pattern of primary protein extract. *New Phytol* 91:705–710
- Petit P, Lallemand R, Savoye D (1983) Purified phytolectin from the lichen *Peltigera canina* var *canina* which binds to the phycobiont cell walls and its use as cytochemical marker *in situ*. *New Phytol* 94:103–110
- Roberts A, Pimentel H, Trapnell C, Pachter L (2011a) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27:2325–2329
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011b) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12:R22
- Rosen S, Bergstrom J, Karlsson KA, Tunlid A (1996) A multispecific saline-soluble lectin from the parasitic fungus *Arthrotrix oligospora*. Similarities in the binding specificities compared with a lectin from the mushroom *Agaricus bisporus*. *Eur J Biochem* 238:830–837
- Sacristan M, Vivas M, Millanes AM, Fontaniella B, Vicente C, Legaz ME (2007) The recognition pattern of green algae by lichenized fungi can be extended to lichens containing a cyanobacterium as photobiont. In: Mendez-Vilas A (ed) *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*
- Sharon N (2007) Lectins: carbohydrate-specific reagents and biological recognition molecules. *J Biol Chem* 282:2753–2764
- Shinohara ML, Loros JJ, Dunlap JC (1998) Glyceraldehyde-3-phosphate dehydrogenase is regulated on a daily basis by the circadian clock. *J Biol Chem* 273:446–452

- Silva MDC, Sá RA, Napoleão TH, Gomes FS, Santos NDL, Albuquerque AC, Xavier HS, Paiva PMQ, Correia MTS, Coelho LCBB (2009) Purified *Cladonia verticillaris* lichen lectin: Insecticidal activity on *Nasutitermes corniger* (Isoptera: Termitidae). *Int Biodeterior Biodegrad* 63:334–340
- Singh RS, Bhari R, Kaur HP (2010) Current trends of lectins from microfungi. *Crit Rev Biotechnol* 1–18
- Singh RS, Bhari R, Kaur HP (2010b) Mushroom lectins: current status and future perspectives. *Crit Rev Biotechnol* 30:99–126
- Sinnemann SJ, Andresson OS, Brown DW, Miao VP (2000) Cloning and heterologous expression of *Solorina crocea* *pyrG*. *Curr Genet* 37:333–338
- Sun H, Zhao CG, Tong X, Qi YP (2003) A lectin with mycelia differentiation and antiphytovirus activities from the edible mushroom *Agrocybe aegerita*. *J Biochem Mol Biol* 36:214–222
- Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515
- Vasta GR (2009) Roles of galectins in infection. *Nat Rev Microbiol* 7:424–438
- Vivas M, Sacristan M, Legaz ME, Vicente C (2010) The cell recognition model in chlorolichens involving a fungal lectin binding to an algal ligand can be extended to cyanolichens. *Plant Biol (Stuttg)* 12:615–621
- Walser PJ, Haebel PW, Kunzler M, Sargent D, Kues U, Aebi M, Ban N (2004) Structure and functional analysis of the fungal galectin CGL2. *Structure* 12:689–702
- Walser PJ, Kues U, Aebi M, Kunzler M (2005) Ligand interactions of the *Coprinopsis cinerea* galectins. *Fungal Genet Biol* 42:293–305
- Walti MA, Walser PJ, Thore S, Grunler A, Bednar M, Kunzler M, Aebi M (2008) Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*. *J Mol Biol* 379:146–159
- Xavier BB, Miao VPW, Jónsson ZO, Andr sson OS (2012) Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: Features and phylogeny . *Fungal Biology* (in press)
- Yagi F, Hiroyama H, Kodama S (2001) *Agrocybe cylindracea* lectin is a member of the galectin family. *Glycoconj J* 18:745–749

PAPER 2

LEC-2, A HIGHLY VARIABLE LECTIN IN THE LICHEN *PELTIGERA MEMBRANACEA*

Sheeba S. Manoharan, Vivian P. W. Miao, and Ólafur. S. Andrésson.
Symbiosis (2012) 58:91–98. DOI 10.1007/s13199-012-0206-y

LEC-2, a highly variable lectin in the lichen *Peltigera membranacea*

Sheeba S. Manoharan · Vivian P. W. Miao ·
Ólafur S. Andrésson

Received: 25 September 2012 / Accepted: 26 November 2012 / Published online: 15 December 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Lectins are a diverse group of carbohydrate binding proteins often involved in cellular interactions. A lectin gene, *lec-2*, was identified in the mycobiont of the lichen *Peltigera membranacea*. Sequencing of *lec-2* open reading frames from 21 individual samples showed an unexpectedly high level of polymorphism in the deduced protein (LEC-2), which was sorted into nine haplotypes based on amino acid sequence. Calculations showed that the rates of nonsynonymous versus synonymous nucleotide substitutions deviated significantly from the null hypothesis of neutrality, indicating strong positive selection. Molecular modeling revealed that most amino acid replacements were around the putative carbohydrate-binding pocket, indicating changes in ligand binding. Lectins have been thought to be involved in the recognition of photobiont partners in lichen symbioses, and the hypothesis that positive selection of LEC-2 is driven by variation in the *Nostoc* photobiont partner was tested by comparing mycobiont LEC-2 haplotypes and photobiont genotypes, as represented by the *rbcLX* region. It was not possible to pair up the two types of marker sequences without conflicts, suggesting that positive selection of LEC-2 was not due to variation in photobiont partners.

Keywords LEC-2 · *Peltigera membranacea* · Polymorphism · Positive selection · Evolution · Lectin

1 Introduction

Lectins are single- or multi-domain glycoproteins capable of binding sugar moieties through specific interactions with carbohydrate recognition domains (Brown et al. 2007). Most lectins are multimeric, consisting of non-covalently associated subunits that give lectins their ability to agglutinate cells or form aggregates with glycoconjugates (Mohr and Pommerening 1985). First discovered in plants, lectins have been found in nearly all eukaryotes. Specific functions of lectins vary substantially but lectins are most often categorized based on their ligand binding specificity. For instance, galectins are able to specifically recognize β -galactosides. Lectins have long been thought to play a role in protection and in symbiosis through recognition of sugar moieties characterizing the surfaces of specific organisms (Sharon and Lis 2004). The necessity to either recognize partners (in order to trigger defense, infection or mutualism) or to escape recognition (to avoid infection or defense mechanisms) is a selective force shaping the evolutionary history of genes in interacting species (De Mita et al. 2006).

Coevolution, the process of reciprocal changes between interacting species during evolution, molds the organization of communities that exist over periods long enough to allow selection of species traits (Thompson 2001). Rapid change driven by positive selection is often detected in genes involved in the coevolutionary process, such as recognition mechanisms in host-pathogen relationships, immune responses, and reproduction processes (Baum et al. 2002; Biswas and Akey 2006; Hughes 2007; Iguchi et al. 2011). Lectins have frequently been implicated in studies of coevolution, and high amino acid

S. S. Manoharan · Ó. S. Andrésson (✉)
Department of Life and Environmental Sciences,
University of Iceland,
101 Reykjavík, Iceland
e-mail: osa@hi.is

S. S. Manoharan
e-mail: ssm5@hi.is

V. P. W. Miao
Department of Microbiology and Immunology,
University of British Columbia,
Vancouver, Canada
e-mail: vmiao@telus.net

diversity reflecting positive selection has been reported in a wide range of organisms including prawns (Ren et al. 2012), amoeboids (Weedall et al. 2011) and oysters (Wu et al. 2011). In corals, it has been shown that lectins change rapidly, possibly due to selective pressure from symbionts (Hayes et al. 2010; Iguchi et al. 2011). Under positive selection, the ratio of nonsynonymous (amino acid-changing) substitutions is higher than that of synonymous (silent) substitutions per site (Hughes and Nei 1988; Yang and Bielawski 2000). In contrast, genes involved in house-keeping and in major morphological and developmental adaptations of individual organisms generally exhibit relatively low levels of amino acid change (Ford 2002).

The lichen *Peltigera membranacea* is an association of two symbionts: a heterotrophic ascomycete (mycobiont), and a photosynthetic cyanobacterium (photobiont) from the genus *Nostoc* (Miao et al. 1997). Lichen mycobionts belonging to the genera *Peltigera* and *Nephroma* appear to show some selectivity in their choice of *Nostoc* strains, but the diversity varies with lichen species (Paulsrud et al. 2000). In the gelatinous *Collembataceae* lichens, reciprocal high co-specificity has been reported as certain species were found to be associated with a single lineage of *Nostoc* (Otálora et al. 2010). *Peltigera* species have served as models in several studies of lectins and some of the extracellular mycobiont lectins distinguish between symbiotic and other cultured *Nostoc* sp., suggesting their involvement in recognition of potential symbiotic partners (Lockhart et al. 1978; Petit et al. 1983; Galun and Kardish 1995; Diaz et al. 2011). Recently *lec-1*, a mycobiont gene encoding a galectin-like protein in *P. membranacea*, was characterized and shown to be differentially expressed in thalli, rhizines and apothecia, tissues of this foliose lichen (Miao et al. 2012). The current study describes the characterization of a second lectin gene, *lec-2*, in *P. membranacea*. We examined whether the corresponding protein LEC-2 is under positive selection, and whether variants of LEC-2 are associated with specific photobiont strains.

2 Materials and methods

2.1 Lichen collection, DNA extraction, amplification and sequencing

The *lec-2* gene was identified in a database of contigs (contiguous sequence blocks) from the on-going *P. membranacea* whole genome sequence (Pmb-WGS) project (unpublished; Xavier et al. 2012). In brief, the Pmb-WGS is comprised of *P. membranacea* metagenomic DNAs extracted from lichen samples collected at Keldnagil, Iceland. DNAs were processed for sequencing at commercial facilities via the Roche 454 and the Illumina/Solexa methodologies, resulting in 1.76

Gb of 454 data and 1.4 Gb of Illumina data that were assembled using automated procedures. For the present study, additional specimens were collected from the same locality, as well as from other sites in Iceland in 2010–2011 (Table 1). DNA was extracted from one lobe of each specimen using methods described previously (Xavier et al. 2012). The lobes from the new Keldnagil specimens were each processed as three to five fragments of ~1 cm² in order to examine intrathalline variability. Vouchers of each specimen were saved in the Andr sson lab herbarium.

Homologs of *lec-2* were amplified by PCR using the primer pairs *lec2-F92* (5'-GTCGTGTCAAATCACTCAAGGTCGG-3') and *lec2-R766* (5'-CCGTAGTCGCCTATATCATCGCA-3'); *rbcLX* (from end of ribulose biphosphate carboxylase gene *L* to start of gene *S*, ~0.8 Kb) was amplified with *CW* (5'-CGTAGCTTCCGGTGGTATCCACGT-3') and *CX* (5'-GGGGCAGGTAAGAAAGGGTTTCGTA-3') (Rudi et al. 1998). Amplification reactions were performed in a final volume of 25 µl containing 1x PCR buffer (Fermentas) (20 mM Tris-HCl (pH8.8 at 25 °C), 10 mM (NH₄)₂SO₄, 10 mM KCl, 1 % (v/v) Triton X-100, 0.1 mg/ml BSA), 2.5 mM MgSO₄, 0.2 mM each dNTP, 0.24 µM primers and 0.2 U *Taq* DNA polymerase. PCR was performed with an initial denaturation step of 2 min at 94 °C, linked to 33 cycles 20 s at 94 °C, 20 s at 55 °C and 50 s at 72 °C, followed by an extension step of 10 min at 72 °C. All *lec-2* amplicons from Keldnagil were digested with *RsaI* and those representing the same lobe were confirmed to have the same restriction pattern (i.e. there was no intrathalline variation) prior to processing one amplicon from each thallus for sequencing. Amplicons were then treated with exonuclease I and alkaline phosphatase (New England BioLabs) and Sanger sequencing of both strands (with above primers) was conducted using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), and analysed on an ABI3100 automated sequencer (Applied Biosystems). DNA sequences were deposited in GenBank as accessions presented in Table 1.

2.2 Identification and characterization of *lec-2*

The *lec-2* gene was identified from Pmb-WGS contigs using the sequence of LEC-1 (accession number JQ899138; Miao et al. 2012) as a query in tblastn (Altschul et al. 1990). The gene was annotated by comparison with known lectin genes and inspection for conserved dinucleotides (GT/AG) at prospective splice sites associated with appropriate frame shifts, and deduced features were confirmed by mapping Illumina RNA-Seq reads from the *P. membranacea* transcriptome sequencing project (Xavier et al. 2012) using the CLC Genomics Workbench version 5.2 (CLC bio, Denmark). A 3D model of LEC-2 was constructed using the SWISS-MODEL Protein Modelling Server (<http://swissmodel.expasy.org/>) (Schwede et al. 2003; Arnold et al. 2006; Kiefer et al. 2009).

Table 1 *Peltigera membranacea* samples used in the study

Location	Geographical coordinates	Voucher	LEC-2 haplotype	<i>lec-2</i> accession	<i>rbcLX</i> accession
Akureyri	65° 40.452' N, 18° 5.550' W	XBB25 (LA-31096)	5	JX876569	JX876596
Bláskógar	64° 17.328' N, 21° 3.303' W	HH139	1	JX876563	JX876597
Botnsdalur	64° 22.691' N, 21° 16.050' W	HH137	4	JX876568	JX876598
Brúarhlöð	64° 15.649' N, 20° 12.783' W	XBB29	7	JX876570	JX876592
Háihver	63° 56.172' N, 19° 10.275' W	XBB60	3	JX876573	JX876603
Hjörleifshöfði	63° 25.52' N, 18° 47.063' W	SSM58	2	JX876572	JX876602
Kaldakvísl	64° 10.644' N, 21° 34.003' W	SSM48	9	JX876574	JX876593
Keldnagil	64° 7.786' N, 21° 46.568' W	XBB13 ^a	9, 10, 11, 1	JX876604-07	-
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO68	1	JX876566	JX876583
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO74	9	JX876575	JX876600
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO82	9	JX876576	JX876601
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO107	9	JX876577	JX876595
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO111	9	JX876578	JX876588
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO113	9	JX876581	JX876586
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO114	4	JX876567	JX876585
Keldnagil	64° 7.786' N, 21° 46.568' W	JEO115	9	JX876580	JX876587
Kelduhverfi	66° 1.367' N, 16° 30.213' W	HH146	1	JX876564	JX876591
Laugardalshellir	64° 13.153' N, 20° 52.905' W	XBB28	6	JX876582	JX876590
Loðmundarfjörður	65° 21.626' N, 13° 51.294' W	HH154	1	JX876565	JX876589
Öskjuhlöð	64° 7.839' N, 21° 55.196' W	JEO70	9	JX876579	JX876584
Skafafell	63° 59.282' N, 16° 53.431' W	HH145	8	JX876571	JX876594
Stóruð	65° 30.874' N, 13° 59.408' W	HH148	1	JX876562	JX876599

^a *Peltigera membranacea* whole genome sequence (Pmb-WGS) project

2.3 Sequence alignment and phylogenetic analyses

Single nucleotide polymorphisms (SNPs) were resolved into haplotypes by determining overlapping Sanger or 454 sequence reads containing identical SNPs (Table 1, Fig. 1). Sequences were aligned using CLUSTALW as implemented in the CLC Genomics Workbench version 5.2 (CLC bio, Denmark). Phylogenetic trees were constructed using maximum likelihood (PhyML 3.0) (Guindon and Gascuel 2003; Guindon et al. 2010) as implemented in Seaview (4.2.12) (Gouy et al. 2010). A model generator was used to identify the best-fit model (Keane et al. 2006) and models TN93 and K80 were used for *lec-2* and *rbcLX* sequences, respectively. Maximum likelihood branch support was determined by statistical analysis using the approximate likelihood-ratio test (Anisimova and Gascuel 2006).

2.4 Test for selection

The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N), and their variances, $\text{Var}(d_S)$ and $\text{Var}(d_N)$ were calculated. The null hypothesis of no selection ($H_0: d_N = d_S$) was tested against the positive selection

hypothesis ($H_1: d_N > d_S$) using the Z test: $Z = (d_N - d_S) / \sqrt{\text{Var}(d_S) + \text{Var}(d_N)}$. This test determines whether d_N/d_S ratios are significantly different from 1 (Nei and Gojobori 1986). Calculations were performed using MEGA version 5 (<http://www.megasoftware.net>) (Tamura et al. 2011).

3 Results

3.1 Identification and characterization of *lec-2*

The sequence of LEC-1 was used to query the *P. membranacea* whole genome sequence (Pmb-WGS) for contigs with additional lectin-like genes. A 3.5 kb contig was identified with a 600 nt open reading frame similar to *lec-1* (Miao et al. 2012) and the gene was designated as *lec-2*. The *lec-2* gene is comprised of a 241 nucleotide (nt) 5' untranslated region (UTR) with one intron (54 nt), two coding regions (472 nt, 8 nt) separated by an intron (64 nt), and a 401 nt 3' UTR. Illumina RNA-Seq data (unpublished) was used to verify the intron junctions. The predicted 17.4 kDa protein of 159 amino acids, LEC-2, was most similar to LEC-1 (GenBank entry JQ899138; 38 % identity) and galectins from the saprotrophic basidiomycete, *Coprinopsis cinerea* (GenBank entries

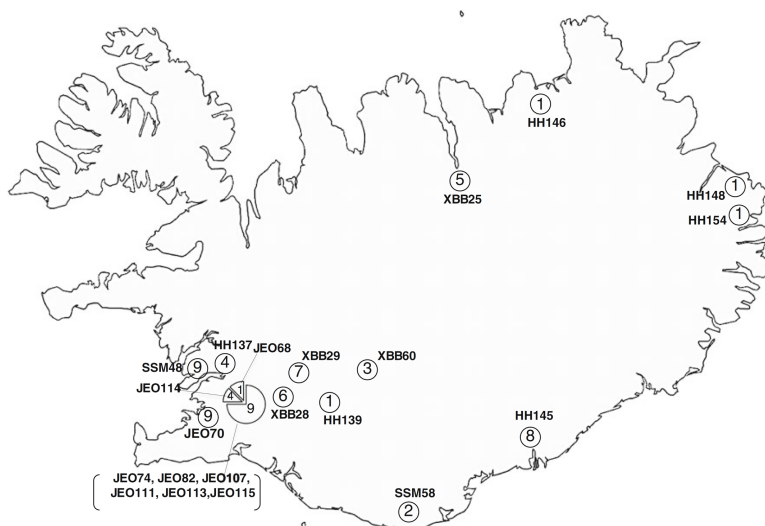


Fig. 1 Sampling sites in Iceland and LEC-2 haplotypes. Approximate location of collection site is indicated by position of circled number presenting LEC-2 haplotype. Geographical coordinates are given in

Table 1. Pie chart in the lower left represents sample site at Keldnagil where three haplotypes were found in different proportions in 2010–2011

XP_001836010.2; 36 % identity and XP_001836012; 31 % identity). A search against proteins with structural models gave the best hit to a galectin from another

saprotrophic, mushroom-forming basidiomycete, *Agrocybe aegerita* (PDB accession, 2ZGP_A, 32 % sequence identity).

		20	40	60	H ⁴⁴	R ⁴⁸	V ⁵⁹ N ⁶¹
* XBB13 [1]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
SSM58 [2]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
XBB60 [3]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
JEO114 [4]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
XBB25 [5]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
XBB28 [6]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
XBB29 [7]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
HH145 [8]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
JEO74 [9]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
* XBB13 [10]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
* XBB13 [11]	MASQWIFLEPTVSPK	LKSTSLDKPFAPGQS	ITFVSTSFDP	IADTTVNLD	SAGFYFTKDG	DVLLSISIRRHQNL	IIFNSR
	W ⁵⁸	E ⁷¹ R ⁷³	100	120	140		
* XBB13 [1]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
SSM58 [2]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
XBB60 [3]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
JEO114 [4]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
XBB25 [5]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
XBB28 [6]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
XBB29 [7]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
HH145 [8]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
JEO74 [9]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
* XBB13 [10]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI
* XBB13 [11]	QGGTWGHEQL	IDLQGVFPGPRA	ITHVTANATKYNIAFNNS	NIHTFTKVI	IQGDP	TVGLHFE	SNKPVFSDPVI

Fig. 2 LEC-2 allelic variation in *P. membranacea*. XBB13 (haplotype 1) is reference; galectin signature residues of the carbohydrate recognition domain are identified at the top (human galectin-1 numbering;

Hirabayashi and Kasai 1988). Amino acid substitutions are indicated by shading. LEC-2 haplotypes are indicated within square brackets and * indicates haplotypes found in the Pmb-WGS project

3.2 Allelic variation in *lec-2* and positive selection

The DNA used for the Pmb-WGS was derived from a mixture of thalli, and when the individual 454 reads were inspected, a high level of variation was found. Four alleles, generating four variants of the deduced protein, LEC-2, were observed and the most common amino acid sequence was designated as haplotype 1 (Table 1, Fig. 2). For comparison, in the same material, the housekeeping genes *gpd-1* (encoding glyceraldehyde-3-phosphate dehydrogenase, accession JQ837250) and *tub-2* (encoding β -tubulin, accession JQ837249) showed no polymorphisms, although the combined length of the coding regions was 3938 nt (Miao et al. 2012). Two of the four LEC-2 haplotypes were again observed when samples of single thalli collected later from the same site were examined individually (Table 1), thus validating the results from whole genome sequencing.

The *lec-2* genes of single lobes collected from 13 additional sites across Iceland were investigated by Sanger sequencing (Table 1). From 21 thalli, a total of 11 protein haplotypes were distinguished (Fig. 2) based on a total of 14 nonsynonymous nucleotide substitutions and one synonymous nucleotide substitution in the coding region of *lec-2*. According to the hypothesis of selective neutrality, the pairwise ratio of d_N/d_S , (number of nonsynonymous substitutions per nonsynonymous site and the number of synonymous substitutions per synonymous site) is equal to 1 (Nei and Gojoberi 1986). The d_N/d_S ratio in the *lec-2* coding region was found to be 5.1, and application of the Z test (Tamura et al. 2011) showed that this deviation from neutrality was highly significant ($p=0.003$).

3.3 Three dimensional (3D) model of LEC-2

A 3D model for *P. membranacea* LEC-2 was generated based on the crystal structure of a galectin from *A. aegerita* (PDB accession 2ZGP_A). Of seven canonical positions associated with ligand binding (human galectin-1 numbering; Hirabayashi and Kasai 1988), LEC-2 showed conservation in five positions and substitutions at two: H⁴⁴S and R⁷³L/S. In this model, amino acid replacements were found mainly on the well-exposed peripheral surface of the protein, rather than in the predominant β -sheet structures (Figs. 2 and 3). About half of the polymorphic sites were on the periphery of the carbohydrate-binding pocket and one-third was in a second fairly localized area.

3.4 Correlation of LEC-2 haplotypes and photobiont *rbclX* genotypes

The association of particular mycobiont and photobiont strains was investigated by examining the correlation between the LEC-2 haplotype and the genotype of the corresponding photobiont. Since the genome of the *Nostoc* is comprised of a

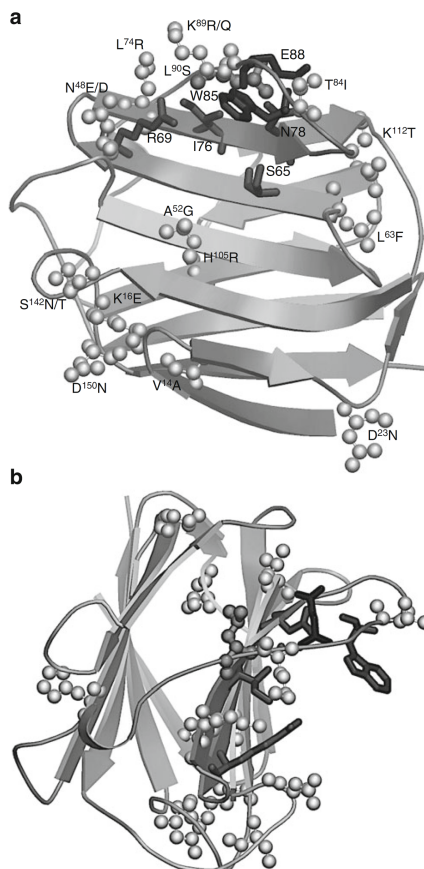


Fig. 3 (a, b) Two views of a 3D model of LEC-2. A ball and stick structure representing the residue of the reference sequence is shown at each polymorphic position; substitutions are indicated by a one-letter amino acid code and galectin signature residues of the carbohydrate recognition domain are represented as cylinders. b) Is rotated 90° clockwise relative to a) and tilted slightly

single chromosome, i.e. one linkage group for all loci, the sequence of one locus such as *rbclX* can be used as a proxy for its strain type, and by inference, its possible lectin ligand. Thus, pairs of sequences covering the coding region of *lec-2* from the mycobiont and the *rbclX* region of the photobiont from each of the 21 individually collected *P. membranacea* thalli were obtained and the sequences for each marker were

aligned and used for constructing phylogenetic trees. The two trees showed different topologies (Fig. 4). Different *Nostoc* strains were found to be associated with a single LEC-2 haplotype; e.g. *Nostoc* strains associated with the most common LEC-2 haplotypes, 1 and 9, were distributed throughout the *rbclX* phylogenetic tree. This suggested a low specificity for the photobiont, with a broad range of photobionts being compatible with some mycobiont LEC-2 haplotypes.

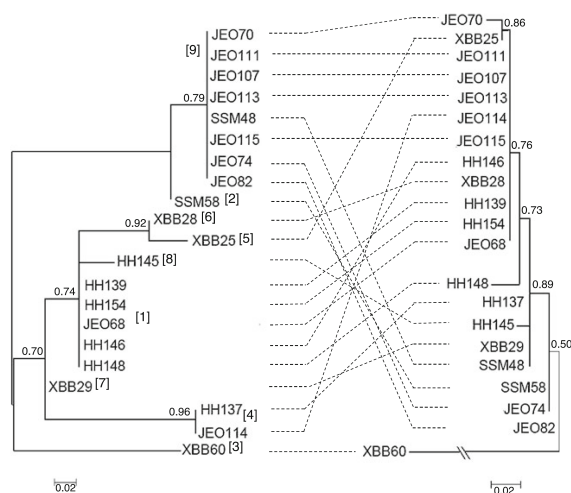
4 Discussion

Lectin-like genes have been identified in various organisms including slime molds (Huh et al. 1998) and fungi (Nowrousian and Cebula 2005; Singh et al. 2010, 2011; Miao et al. 2012). In this study, we describe the *lec-2* gene from the mycobiont of the lichen *P. membranacea*. The high sequencing coverage in the Pmb-WGS and typical eukaryotic sequence features, including the presence of an intron and fungal splicing signals, support the conclusion that *lec-2* is a mycobiont gene. Alignment of LEC-2 with LEC-1 (Miao et al. 2012) revealed 38 % identity and the primary structure of the LEC-2 sequence showed similarity to fungal galectins CGL1 and CGL2 from *C. cinerea*, both involved in fruiting body development (Walser et al. 2005). The structural model of LEC-2, based on a galectin from *A. aegerita*, showed galectin-type ligand coordinating residues around a presumptive carbohydrate-binding pocket (Fig. 3).

In contrast to the total lack of SNPs in the previously described *lec-1* (Miao et al. 2012), considerable sequence

diversity was found in the coding region of *lec-2* when data from the Pmb-WGS project was examined. While intra-lobe samples consistently yielded a single restriction pattern, a total of 14 amino acid substitutions was found in 21 samples from 14 geographical locations. This number is high compared with what has been described in other taxa of lichen-forming fungi (Werth 2010), given that we examined only 21 thalli, suggesting that there may be strong selection for changes in the functional properties of the LEC-2 protein. Modeling of the LEC-2 structure showed that the substitutions occurred primarily at well-exposed sites on the peripheral surface of the protein (Fig. 3). Amino acids characteristic for the galectin carbohydrate recognition domain R⁴⁸, N⁶¹, W⁶⁸, E⁷¹ (Fig. 2) were found near the proposed sugar-binding pocket, but substitutions in the carbohydrate recognition domain-motif were identified at positions H⁴⁴S, V⁵⁹I and R⁷³L/S. Replacements at H⁴⁴ and R⁷³ have been observed in a galectin found in the mycorrhizal basidiomycete *Laccaria amethystina* (Lyimo et al. 2011), while the conservative replacement V⁵⁹I has been reported in various galectins, including that of *A. aegerita* (Yagi et al. 2001). In the galectin from *C. cinerea*, a W⁶⁸R substitution in the carbohydrate recognition domain is known to be critical in determining chitoooligosaccharide versus lactose binding (Wälti et al. 2008), while in a galectin from *Agrocybe cylindracea*, residues upstream of the carbohydrate recognition domain motif were influential in determining sialic acid binding (Ban et al. 2005). In the galectin of *A. aegerita*, substitutions H⁵⁹A and R⁶³H affected the lactose binding ability (Yang et al. 2009). Thus, substitutions at or close to the galectin carbohydrate recognition domain in LEC-2 are likely to alter

Fig. 4 Phylogenetic trees of LEC-2 haplotype sequences (left) and *rbclX* sequences (right). Bar indicates substitutions per site. Numbers at the branch points show approximate likelihood ratios. Haplotype numbers (square brackets) are shown for each clade. Dotted lines connect sequences from the same sample



the specificity of ligand binding, possibly enabling discrimination of surface sugars on an interacting organism. Additionally, amino acid substitutions at the surface of the protein may alter multimerization, often involved in effective ligand binding and pattern recognition (Sharon and Lis 2004).

Comparison of the relative frequencies of synonymous and nonsynonymous nucleotide substitutions indicated significant positive selection in LEC-2. The selection is most likely due to an important interaction with a factor that shows clear variation, ultimately expressed in differences in ligand coordinating residues in the carbohydrate recognition domain. This driving force for selection on *lec-2* is unlikely to be ascribed to physical factors such as climate or niche, as multiple haplotypes exist within a topographically homogeneous sampling site (Keldnagil), and the same haplotype can be found at geographically distant sites representing both coastal and highland habitats. It appears more likely that the selective response is due to biotic factors associated with symbiont partners, pathogens or parasites.

The hypothesis that evolution of LEC-2 is driven by interaction with different strains of *Nostoc* photobionts was tested by comparing the mycobiont LEC-2 haplotypes and the photobiont-derived *rbcLX* sequences. For genotyping of *Nostoc* symbionts, we chose the *rbcLX* sequence marker, since it has been shown to be more suitable for *Nostoc* typing than 16S rDNA and tRNA^{Leu} (intron) sequences (Otálora et al. 2010). We found no significant correspondence between the *Nostoc* genotypes and the mycobiont LEC-2 haplotypes. Thus, LEC-2 does not seem to be a determinant of photobiont partner choice in *P. membranacea*. As the gene(s) coding for the LEC-2 ligand are unknown, there is uncertainty about their linkage to *rbcLX*, but we assumed linkage within each *Nostoc* strain, as a low degree of recombination is generally observed in clonal bacterial populations (Didelot and Maiden 2010).

Future experimental determination of the nature of the LEC-2 ligand(s) and various LEC-2 haplotypes may help shed light on the interaction of LEC-2 and its ligand, and the role it plays in the biology of *P. membranacea*. Further research should also focus on the distribution of *lec-2* homologs in the genus *Peltigera* as well as the origin and divergence of *lec-2* in *P. membranacea*.

Acknowledgements We thank Starri Heiðmarsson for providing sample LA-31096, Basil Britto Xavier, Juho Rinne and Helga Hjartardóttir for DNA samples and DNA isolations, Eva Ösp Björnsdóttir for help with PCR and Silke Werth for useful discussion.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55:539–552
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201
- Ban M, Yoon HJ, Demirkan E, Utsumi S, Mikami B, Yagi F (2005) Structural basis of a fungal galectin from *Agroclype cylindracea* for recognizing sialoconjugate. *J Mol Biol* 351:695–706
- Baum J, Ward RH, Conway DJ (2002) Natural selection on the erythrocyte surface. *Mol Biol Evol* 19:223–229
- Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends Genet* 22:437–446
- Brown AC, Harrison LM, Kapulkin W, Jones BF, Sinha A, Savage A, Villalon N, Cappello M (2007) Molecular cloning and characterization of a C-type lectin from *Ancylostoma ceylanicum*: evidence for a role in hookworm reproductive physiology. *Mol Biochem Parasitol* 151:141–147
- De Mita S, Santoni S, Hochu I, Ronfort J, Bataillon T (2006) Molecular evolution and positive selection of the symbiotic gene NORK in *Medicago truncatula*. *J Mol Evol* 62:234–244
- Díaz EM, Vicente-Manzanares M, Sacristan M, Vicente C, Legaz ME (2011) Fungal lectin of *Peltigera canina* induces chemotropism of compatible *Nostoc* cells by constriction-relaxation pulses of cyanobiont cytoskeleton. *Plant Signal Behav* 6:1525–1536
- Didelot X, Maiden MC (2010) Impact of recombination on bacterial evolution. *Trends Microbiol* 18:315–322
- Ford MJ (2002) Applications of selective neutrality tests to molecular ecology. *Mol Ecol* 11:1245–1262
- Galun M, Kardish N (1995) Lectins as determinants of symbiotic specificity in lichens. *Cryptogamic Botany* 5:144–148
- Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Hayes ML, Eytan RI, Hellberg ME (2010) High amino acid diversity and positive selection at a putative coral immunity gene (*tachylectin-2*). *BMC Evol Biol* 10:150
- Hirabayashi J, Kasai K (1988) Complete amino acid sequence of a beta-galactoside-binding lectin from human placenta. *J Biochem* 104:1–4
- Hughes AL (2007) Looking for Darwin in all the wrong places: The misguided quest for positive selection at the nucleotide sequence level. *Heredity* 99:364–373
- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature* 335:167–170
- Huh CG, Aldrich J, Mottahedeh J, Kwon H, Johnson C, Marsh R (1998) Cloning and characterization of *Physarum polycephalum* tectonins. Homologues of Limulus lectin L-6. *J Biol Chem* 273:6565–6574
- Iguchi A, Shinzato C, Forêt S, Miller DJ (2011) Identification of fast-evolving genes in the scleractinian coral *Acropora* using comparative EST analysis. *PLoS One* 6:e20140
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection

- and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* 37:D387–D392
- Lockhart CM, Rowell P, Stewart WDP (1978) Phytohaemagglutinins from the nitrogen-fixing lichens *Peltigera canina* and *P. polydactyla*. *FEMS Microbiology Letters* 3:127–130
- Lyimo B, Yagi F, Minami Y (2011) Primary structure and specificity of a new member of galectin family from the amethyst deceiver mushroom *Laccaria amethystina*. *Biosci Biotechnol Biochem* 75:62–69
- Miao VPW, Manoharan SS, Snæbjarnarson V, Andr sson  S (2012) Expression of *lec-1*, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea*. *Symbiosis* 57:23–31
- Miao VPW, Rabenau A, Lee A (1997) Cultural and molecular characterization of photobionts of *Peltigera membranacea*. *Lichenologist* 29:571–586
- Mohr P, Pommerening K (1985) Affinity chromatography—practical and theoretical aspects. Marcel Dekker, Inc., New York
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Nowrousian M, Cebula P (2005) The gene for a lectin-like protein is transcriptionally activated during sexual development, but is not essential for fruiting body formation in the filamentous fungus *Sordaria macrospora*. *BMC Microbiology* 5:64
- Ot lora MAG, Mart nez I, O’Brien H, Molina MC, Arag n G, Lutzoni F (2010) Multiple origins of high reciprocal symbiotic specificity at an intercontinental spatial scale among gelatinous lichens (Collemales, Lecanoromycetes). *Mol Phylogenet Evol* 56:1089–1095
- Paulsrud PER, Rikkinen J, Lindblad P (2000) Spatial patterns of photobiont diversity in some *Nostoc*-containing lichens. *New Phytologist* 146:291–299
- Petit P, Lallemand R, Savoye D (1983) Purified phytolectin from the lichen *Peltigera canina* var *canina* which binds to the phycobiont cell walls and its use as cytochemical marker *in situ*. *New Phytologist* 94:103–110
- Ren Q, Li M, Du J, Zhang CY, Wang W (2012) Immune response of four dual-CRD C-type lectins to microbial challenges in giant freshwater prawn *Macrobrachium rosenbergii*. *Fish Shellfish Immunol* 33:155–167
- Rudi K, Skulberg OM, Jakobsen KS (1998) Evolution of cyanobacteria by exchange of genetic material among phylogenetically related strains. *J Bacteriol* 180:3453–3461
- Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385
- Sharon N, Lis H (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* 14:53R–62R
- Singh RS, Bhari R, Kaur HP (2010) Mushroom lectins: current status and future perspectives. *Crit Rev Biotechnol* 30:99–126
- Singh RS, Bhari R, Kaur HP (2011) Current trends of lectins from microfungi. *Crit Rev Biotechnol* 31:193–210
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Thompson JN (2001) Coevolution. in: Encyclopedia of Life Sciences, London, Nature publishing group
- Walser PJ, K es U, Aebi M, K nzler M (2005) Ligand interactions of the *Coprinopsis cinerea* galectins. *Fungal Genet Biol* 42:293–305
- W ltsi MA, Walser PJ, Thore S, Gr nler A, Bednar M, K nzler M, Aebi M (2008) Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*. *J Mol Biol* 379:146–159
- Weedall GD, Sherrington J, Paterson S, Hall N (2011) Evidence of gene conversion in genes encoding the Gal/GalNAc lectin complex of *Entamoeba*. *PLoS Negl Trop Dis* 5:e1209
- Werth S (2010) Population genetics of lichen-forming fungi—a review. *Lichenologist* 42:499–519
- Wu Q, Li L, Zhang GF (2011) *Crassostrea angulata* bindin gene and the divergence of fucose-binding lectin repeats among three species of *Crassostrea*. *Mar Biotechnol* 13:327–335
- Xavier BB, Miao VP, J nsson ZO, Andr sson  S (2012) Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: Features and phylogeny. *Fungal Biol* 116:802–814
- Yagi F, Hiroshima H, Kodama S (2001) *Agrocybe cylindracea* lectin is a member of the galectin family. *Glycoconj J* 18:745–749
- Yang N, Li DF, Feng L, Xiang Y, Liu W, Sun H, Wang DC (2009) Structural basis for the tumor cell apoptosis-inducing activity of an antitumor lectin from the edible mushroom *Agrocybe aegerita*. *J Mol Biol* 387:694–705
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503

CHAPTER 5

***PELTIGERA ISLANDICA*, A NEW CYANOLICHEN SPECIES IN SECTION *PELTIGERA* ('*P. CANINA* GROUP')**

PAPER 3

***PELTIGERA ISLANDICA*, A NEW CYANOLICHEN SPECIES IN SECTION *PELTIGERA* ('*P. CANINA* GROUP')**

Manoharan–Basil SS, Miadlikowska J, Goward T, Andrésson OS and Miao VPW. *The Lichenologist* 48(5): 451–467 (2016).
doi:10.1017/S002

Peltigera islandica, a new cyanolichen species in section *Peltigera* ('*P. canina* group')

Sheeba S. MANOHARAN-BASIL, Jolanta MIADLIKOWSKA, Trevor GOWARD,
Ólafur S. ANDRÉSSON and Vivian P. W. MIAO

Abstract: A new cyanolichen, *Peltigera islandica* sp. nov. in the section *Peltigera* ('*P. canina* group') is described from Iceland. This species is similar in general appearance to *P. rufescens* and *P. membranacea*, but may be recognized by its downturned lobe tips and narrow lobes, respectively. Most thalli are bright emerald green in colour when moist, although a dark khaki green colourmorph is also documented. Monophyly of *P. islandica* s. lat. (i.e. including *P. sp.* A *sensu* O'Brien *et al.*, from Canada) is significantly supported based on ITS sequences and corroborated by molecular synapomorphy (absence of the ITS1 hypervariable region). Analysis of the *rbcLX* locus indicates the cyanobiont of *P. islandica* (*Nostoc* sp.) comprises strains belonging to a pool of Icelandic genotypes, some of which are present in other *Peltigera* species, including *P. "neorufescens"*, another taxon new to Iceland collected during this study. Association with photobionts that are shared by other local species suggests *P. islandica* may be well established in Iceland, but a review of herbarium collections as well as broader field surveys are needed to better characterize its geographical distribution.

Key words: colourmorph, Iceland, lichen community, *Nostoc*, *Peltigera* "*neorufescens*", *rbcLX*

Accepted for publication 6 June 2016

Introduction

Peltigera is a widely distributed genus of primarily muscicolous and terricolous macrolichens with more than 90 species recognized worldwide (Goward *et al.* 1995; Martinez *et al.* 2003; Vitikainen 2006; Kirk *et al.* 2008; Han *et al.* 2013, 2015). Most of them are bipartite symbioses, with one mycobiont and a *Nostoc* cyanobacterial photobiont. These species have thalli that range in colour from dark green to bluish grey and brown when moist and are grouped into five sections: *Horizontales*, *Hydrothyriae*,

Peltigera, *Polydactylon*, and *Retifoveatae* (Miadlikowska & Lutzoni 2000). Thalli of different colours within a species (colourmorphs) are notable in some bipartite taxa, and in one case this has been correlated with different strains of the *Nostoc* photobiont (Miao *et al.* 1997). The remaining members of the genus are classified in sections *Chloropeltigera*, *Peltidea* and *Phlebia*. With two exceptions (*P. malacea* and *P. frippii*), the members of those three sections are tripartite symbioses, with one mycobiont and two photobionts, and are 'emerald' or 'grass' green. These consort with a green alga, *Coccomyxa*, as the main photobiont in the thallus while a secondary *Nostoc* photobiont is compartmentalized in external cephalodia in the form of bluish or brownish 'warts'.

At the genus level, *Peltigera* are among the most recognizable of terrestrial macrolichens and easily distinguished from others owing to their broad lobes and veined, ecorticate lower surfaces. However, identification of some species may be difficult due in part to intrinsic variation and high phenotypic

S. S. Manoharan-Basil and Ó. S. Andrésson: Department of Life and Environmental Sciences, University of Iceland, 101 Reykjavik, Iceland.

J. Miadlikowska: Department of Biology, Duke University, Durham, NC 27708-90338, USA.

T. Goward: UBC Herbarium, Beaty Museum, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

V. P. W. Miao (corresponding author): Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada. Email: vmiao@telus.net

plasticity (Vitikainen 1994; Goward *et al.* 1995; Goffinet & Bayer 1997; Miadlikowska & Lutzoni 2000, 2004; Miadlikowska *et al.* 2003; O'Brien *et al.* 2009; Sérusiaux *et al.* 2009). Cryptic species and species complexes present special challenges, and previously unrecognized taxa may be discovered only after molecular data have been examined. Examples include undescribed lichens from western North America reported by Miadlikowska *et al.* (2003) in section *Peltigera*, viz. *P. "neorufescens"*, *P. "fuscopraetextata"*, and *P. "neocanina"*. More recently, O'Brien *et al.* (2009), working with material from British Columbia, Canada, suggested yet another potential undescribed species in the same section, referred to simply as *P. sp. A*.

In 2010, the first author collected an emerald green but non-cephalodiate *Peltigera* specimen at Öskjuhlid, a hill in Reykjavík in south-western Iceland. The specimen did not correspond to any of the 22 *Peltigera* species reported from Iceland (Kristinsson & Heiðmarsson 2009). Analysis of the nuclear ribosomal internal transcribed spacer region (ITS) revealed a match to *P. sp. A sensu* O'Brien *et al.* (2009). In light of this finding, Öskjuhlid was revisited, mainly in 2013, to obtain additional specimens and to more closely examine the community of lichens in this habitat.

The main purposes of this contribution are to formally describe *Peltigera islandica* sp. nov., based on collections from Iceland, with an updated phylogeny of the mycobiont, and to characterize its photobiont, particularly in relation to the *Peltigera* community at Öskjuhlid. Additionally, *P. "neorufescens"*, a second undescribed taxon collected during fieldwork in 2013, is reported as new to Iceland.

Materials and Methods

Sample collection and identification

Peltigera islandica was collected from four localities in south-west Iceland (Fig. 1A, Table 1): Öskjuhlid (elevation 40 m a.s.l. 64-131683°N, 21-920017°W), Thverás (64-104183°N, 21-780667°W), Thjórsá (64-03603°N, 20-3152°W), and near Hrauneyjafoss (2 sites: 64-16659°N, 19-1536°W and 64-17036°N, 19-1296°W). At Öskjuhlid, the area near where the first specimen, accession SSM55,

was found in 2010 was marked off in 1.3 m × 0.5 m sectors (Fig. 1B) to support more systematic collecting, as well as to facilitate description and recording of the spatial relationship among samples. In addition to *P. islandica*, representatives of every apparent morphotype in each sector were also collected.

Samples were sorted, cleaned of field debris and soil, and examined under a hand lens or a dissecting microscope for identification. Fresh thalli of *P. islandica* were hand-sectioned using a sterile razor blade and examined through a compound microscope using autofluorescence (Leica model DM2500) and phase contrast (Nikon model Labophot Trinocular Microscope) microscopy, and standard C, K, KC, and PD spot tests were performed on the cortex and medulla (Orange *et al.* 2001). After identification, a piece of each specimen was used for DNA extraction (to corroborate identifications, see below) and the rest was saved for herbarium vouchers. Except for P4 and P14 (Miadlikowska & Lutzoni s. n.) deposited at Duke University herbarium (DUKE), vouchers were saved in the Andrésson laboratory at the University of Iceland, with a portion of some accessions (*P. islandica* SSM55, SSM63, SSM165, SSM344, SSM348, SSM355, SW329, SW330, SW331, OSA332, and "*P. neorufescens*" SSM347, SSM359) also deposited at the Icelandic Institute of Natural History (AMNH).

Isolation and culture of cyanobacterial photobionts

Fresh, washed thallus fragments of accession SSM165 were macerated and spread under sterile conditions on BG11₀ plates and incubated under low light intensities at 25 °C. After three weeks, the cyanobacteria that grew away from the mycobiont were transferred to fresh plates using a sterile needle under a dissecting microscope, and purified cultures were maintained on BG11₀ (Hainfried & Schenk 2013).

Isolation of DNA, PCR amplification and sequencing

Total genomic DNA was isolated from thalli as described previously (Sinnemann *et al.* 2000) for use as PCR templates for mycobiont and photobiont markers. DNA was prepared from replicate portions ('a' and 'b') of accessions SSM63, SSM348 and SSM356 to assess potential genetic heterogeneity of the thallus. For PCR from cultured photobionts, a lysate was first made (a few cells transferred from agar plates into 5% Chelex (Sigma), heated to 100 °C for 2 min, and clarified by centrifugation), and an aliquot of the supernatant was used as the template for PCR.

The mycobiont ITS region was amplified with primers ITS1 and ITS4 (White *et al.* 1990), while part of the RNA polymerase II largest subunit (*RPB1*) and β-tubulin (*tub-2*) loci were amplified using newly designed and/or modified primers: forward R1 (5'-AGAGTGTCCTCGGG CATTTTGGTCA-3') and reverse R2 (5'-TGGCCGGC AATCTCATTATCCATGTA-3') based on Stiller & Hall

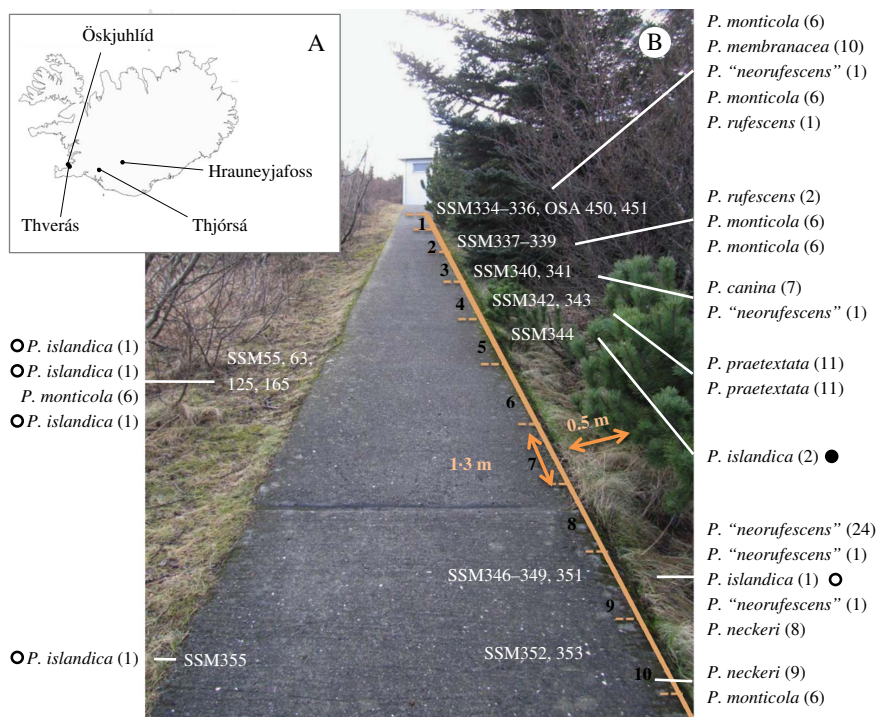


FIG. 1. Collection sites for *Peltigera islandica*. A, location of collection sites in Iceland; B, collection site at Öskjuhlíð. Cover vegetation (April 2013) and sectors along the path where samples were collected are shown. The order of voucher numbers corresponds to the order of the species listed. Appended to each name is the corresponding *rbclX* allele (Table 1). ○ = emerald green; ● = dark green colourmorphs of *P. islandica*. In colour online.

(1997) and Matheny *et al.* (2002), and forward T1 (5'-ACCTTCGTGCGGATACACACT-3') and reverse T2 (5'-ACCCTCAGTGTARTGACCCTTG GC-3') based on Glass & Donaldson (1995). The photobiont *rbclX* locus was amplified using primers CW and CX (Rudi *et al.* 1998), while the 16S rDNA gene was amplified from SSM55 using primer 8F (Turner *et al.* 1999) and a new 16S rDNA reverse primer, PiR (5'-AAAGGAGGTGATCCAGCCAC-3'). Reactions were performed in 25 µl containing 1 × PCR buffer (Fermentas), 2.5 mM MgSO₄, 0.2 mM each dNTP, 0.24 µM primers and 0.2 U *Taq* DNA polymerase, except for *tub-2* for which 1 × Eppendorf Master Mix was used. Amplifications were performed with an initial heating step of 94 °C for 2 min (5 min when using lysates) and 33 cycles of 94 °C for 20 s, 55 °C for 20 s and 72 °C for 50 s (90 s when amplifying *tub-2*). PCR products were treated with exonuclease I and then alkaline phosphatase (ExoSAP, New England BioLabs) to degrade single stranded DNA and inactivate remaining nucleotides, and subjected to cycle

sequencing using ABI Big Dye Terminator v3.1. PCR amplification primers were used for sequencing, except for the 16S amplicon for which internal primers from the literature (e.g. 515F, 519R; Turner *et al.* 1999) were used to complete primer walking. Reaction products were run on an ABI3500XL genetic analyzer, and automated base calls were manually confirmed after inspection of chromatograms. Sequences, including 41 ITS, 19 *RPB1*, 15 *tub-2* and 35 *rbclX*, were compared by BLAST searches in public databases to corroborate species identification, then annotated and deposited in GenBank (Table 1) as part of this study ahead of manuscript completion. Comparison to public databases was last conducted on 24 April 2016.

Sequence alignment and phylogenetic analyses

The ITS dataset used for phylogenetic analysis of the mycobionts included 19 newly generated sequences

TABLE 1. Specimen information and GenBank accession numbers for the species of *Peltigera* from Iceland used in this study.

Taxon	Location	Voucher	ITS	GenBank Accession Number				<i>rbcLX</i> type
				<i>tub-2</i>	<i>RPB1</i>	<i>rbcLX</i>		
<i>P. canina</i> s. lat.	Vaglaskógur	SSM54	KC139756	KC139768	KC139764			
<i>P. canina</i> s. lat.	Hrafnhólar	SSM56	KC139758	KC139770	KC139766			
<i>P. canina</i> s. lat.	Öskjuhlid	SSM340	KJ413197			KJ413212		7
<i>P. canina</i> s. lat.	Arnessýsla	P4	KU954062					
<i>P. canina</i> s. str.	Akureyri	P14	KU954063					
<i>P. islandica</i>	Öskjuhlid	SSM55	KC139757	KC139769	KC139765	KC139773		1
<i>P. islandica</i>	Öskjuhlid	SSM63a	KC139759	KC139771	KC139767	KC139774		1
<i>P. islandica</i>	Öskjuhlid	SSM165	KC139761	KC139772		KC139776		1
<i>P. islandica</i>	Öskjuhlid	SSM344	KJ413241	KJ413191	KJ413228	KJ413208		2
<i>P. islandica</i>	Öskjuhlid	SSM348a	KJ413242	KJ413193	KJ413229	KJ413199		1
<i>P. islandica</i>	Öskjuhlid	SSM355	KJ413244	KJ413192	KJ413230	KJ413198		1
<i>P. islandica</i>	Hrauneyjafoss	SW329	KJ413240	KJ413194	KJ413233	KJ413216		1
<i>P. islandica</i>	Hrauneyjafoss	SW330	KJ413238	KJ413195	KJ413232	KJ413217		3
<i>P. islandica</i>	Thjórsá	SW331	KJ413239	KJ413196	KJ413231	KJ413218		4
<i>P. islandica</i>	Thverás	OSA332*	KJ413245	KJ413189	KJ413227	KJ413219		5
<i>P. membranacea</i>	Öskjuhlid	SSM335				KJ413213		10
<i>P. monticola</i>	Öskjuhlid	SSM125	KC139753		KC139752	KT956444		6
<i>P. monticola</i>	Öskjuhlid	SSM334	KJ413224			KJ413204		6
<i>P. monticola</i>	Öskjuhlid	SSM338	KJ413222			KJ413201		6
<i>P. monticola</i>	Öskjuhlid	SSM339	KJ413223			KJ413202		6
<i>P. monticola</i>	Öskjuhlid	SSM353	KJ413221			KJ413203		6
<i>P. monticola</i>	Öskjuhlid	OSA450	KT948755			KT956445		6
<i>P. neckeri</i>	Öskjuhlid	SSM351	KJ413250			KJ413210		8
<i>P. neckeri</i>	Öskjuhlid	SSM352	KJ413249			KJ413209		9
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM076	KT861423			KU248941		1
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM077	KT861422 [†]					
<i>P. "neorufescens"</i>	Öskjuhlid	SSM336	KJ413248	KJ413237		KJ413206		1
<i>P. "neorufescens"</i>	Öskjuhlid	SSM341	KJ413247			KJ413205		1
<i>P. "neorufescens"</i>	Öskjuhlid	SSM346	KT861425			KT875357		24
<i>P. "neorufescens"</i>	Öskjuhlid	SSM347	KJ413246	KJ413236	KJ413235	KJ413207		1
<i>P. "neorufescens"</i>	Öskjuhlid	SSM349	KT861424			KT875354		1
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM356a	KJ616386	KJ616379	KJ616390	KJ616376		1
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM358	KJ616385	KJ616383	KJ616391	KJ616375		1
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM359	KJ616388	KJ616381	KJ616389	KJ616378		1
<i>P. "neorufescens"</i>	Mosfellsbaer	SSM360	KJ616384	KJ616382		KJ616374		1
<i>P. praetextata</i>	Öskjuhlid	SSM342	KJ413225			KJ413214		11
<i>P. praetextata</i>	Öskjuhlid	SSM343	KJ413226			KJ413215		11
<i>P. praetextata</i>	Thríhyrningur	DB5097	KC139749					
<i>P. praetextata</i>	Seljalandsfoss	DB5098	KC139750					
<i>P. rufescens</i>	Kaldaklof	SSM044				KU168039		4
<i>P. rufescens</i>	Öskjuhlid	SSM337	KJ413220			KJ413211		2
<i>P. rufescens</i>	Öskjuhlid	OSA451	KT948756			KT956443		1

* *P. islandica* OSA392, collected in the summer of 2015, is not listed separately above because it was part of the same thallus sampled and accessioned as OSA332 in 2013. The ITS sequence of OSA392 was confirmed as identical to that of OSA332.

[†] While other *P. "neorufescens"* are identical to each other and similar to AY257910, SSM077 is similar to AY257916.

(Table 1) and 22 sequences from GenBank. The latter represented *P. sp. A* and seven species closely related to *P. sp. A* (all members of section *Peltigera* ("*P. canina* group") including three undescribed taxa, *P. "boreorufescens"*, *P. "fuscopraetextata"* and *P. "pallidorufescens"*), selected based on published phylogenies (Miadlikowska *et al.*

2003; O'Brien *et al.* 2009; Sérusiaux *et al.* 2009) and GenBank BLAST results (highest similarity to *P. islandica*), and two outgroup species (*P. tereziana* and *P. degenii*). Identical haplotypes were collapsed and a total of 22 ITS sequences were aligned manually with MacClade 4.07 (Maddison & Maddison 2003).

Ambiguously aligned regions, *sensu* Lutzoni *et al.* (2000), were delimited manually (105 of a total of 619 characters) and excluded. Phylogenetic analyses of the remaining 514 characters were performed using the maximum likelihood (ML) optimization criterion as implemented in RAxML (Stamatakis 2006; Stamatakis *et al.* 2008) and Bayesian analysis with BALi-Phy (Suchard & Redelings 2006). A maximum likelihood search for the most likely tree (1000 replicates) and bootstrap analysis (1000 bootstrap pseudoreplicates) were performed using RAxMLHPC-MPI-SSE3 as implemented in the Mobyli SNAP Workbench version 1.0.5, a portal for evolutionary and population genetics analyses (North Carolina State University online facilities, Monacell & Carbone (2014)) with the six parameter substitution model and gamma distribution approximated with four categories (GTRGAMMA; Rodriguez *et al.* 1990) applied independently to the ITS1 and ITS2 regions (one partition) and 5.8S region (second partition). The nexus file for the ML analysis was deposited in TreeBASE (<http://purl.org/phylo/treebase/phylo/study/19185>). Bayesian analysis was performed on a non-censored alignment sampled from the joint posterior distribution of phylogeny, alignment, and model parameter; for detailed information see Redelings & Suchard (2005), Gaya *et al.* (2011) and Sliwa *et al.* (2012). The alignment for 5.8S sequences was held constant, while the alignment for ITS1 and ITS2 was allowed to vary through the Markov chain Monte Carlo (MCMC) process. The substitution model parameters were estimated for the ITS1 and ITS2 regions using the Akaike Information Criterion (AIC) as implemented in MrModeltest 2.3 (Nylander 2004). A six-rate parameter model for nucleotide substitutions (GTR + INV) with a proportion of invariable sites was used for ITS1, and the HKY model with a gamma distribution (Hasegawa *et al.* 1985) was implemented for 5.8S and ITS2. The relative substitution rate for each partition was allowed to vary independently. The alignment algorithm for the ITS1 and ITS2 partitions followed the RS07 indel model (Redelings & Suchard 2007) with the same parameters; 100 000 iterations were sampled from each of eight MCMC chains, and the first 45 000 iterations were discarded as burn-in (assessed using the web tool Are We There Yet (AWTY; Wilgenbusch *et al.* 2004) as implemented on the website http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php).

The *Nostoc* locus used for determining photobiont haplotypes designated in Table 1 corresponded to *rbclX* nucleotide positions 13–794 of GenBank Z94891 (Rudi *et al.* 1998) and was processed using alignment tools and ClustalW as implemented in MacVector 12.0.02. This locus includes sequences encoding the last 82 amino acids of the RUBISCO large subunit (*rbclL*), a putative chaperone gene (*rbclX*) and two intergenic spacers (Li & Tabita 1997). For phylogenetic analysis, the dataset of 295 OTUs comprised published *rbclX* sequences of free-living and symbiotic strains from *Nostoc* clades I and II (Otalora *et al.* 2010; O'Brien *et al.* 2013; Miadlikowska *et al.* 2014a), symbionts from *Gunnera magellanica*, as well as published (Manoharan *et al.* 2012) or new (this study) sequences from *P. islandica* and other *Peltigera* species

from Iceland. In this more diverse dataset, both spacers were aligned poorly (positional homology could not be determined unambiguously) and therefore were removed, leaving 636 coding characters. RAxML-HPC2 v. 7.2.8, as implemented on the CIPRES portal (Miller *et al.* 2010) was used for analysis. Optimal tree and bootstrap searches were conducted with the rapid hill-climbing algorithm for 1000 replicates with the GTRGAMMA substitution model estimated independently for three partitions corresponding to the first, second, and third codon positions of the *rbclX*.

Results

Collection of *P. islandica* sp. nov. and Öskjuhlíd *Peltigera* community

Öskjuhlíd, the hill where most of the material in this study was collected, was cleared in the 1940s for installation of above-ground municipal hot water tanks. After these were demolished in the 1960s, recreational trails were laid on top of the buried pipes, and the adjacent area has grown up to various herbs and shrubs, as well as to downy birch (*Betula pubescens*), Sitka spruce (*Picea sitchensis*), and mountain pine (*Pinus mugo*), introduced through afforestation (Sigurðsson 1993). Terrestrial lichens (*Cladonia* spp., *Peltigera membranacea* and crustose species) were sparse in the open, grassy area NE of the study site, and somewhat more abundant in the wooded area towards the SW. Except for infrequent, scattered occurrences of *Cladonia* species, lichens were generally absent next to paths, but the first specimen, SSM55, was found in 2010 in a small area with an unusually large number of *Peltigera* thalli. Most of this *Peltigera* community was distributed on the SW side of the path, with some sectors supporting one to several thalli or patches and other sectors having none (e.g. sector 1 vs 7, Fig. 1B). Thus, while the focus of the 2013 fieldwork was to obtain more specimens such as SSM55, samples of the other *Peltigera* in the vicinity were also collected in order to complement the morphological description of the new lichen with information on its ecology.

In total, 25 samples (including SSM55) representing eight morphospecies were collected (Table 1). Six were previously reported

from Iceland: *P. canina*, *P. membranacea*, *P. monticola*, *P. praetextata* and *P. rufescens* in section *Peltigera*, and *P. neckeri* in section *Horizontales* (Kristinsson & Heiðmarsson 2009). The seventh species is currently informally designated *P. "neorufescens"*, and represents an undescribed member of the *P. rufescens* complex first recognized by Miadlikowska *et al.* (2003) through ITS and ribosomal gene sequence data. This species was found at one other location in Iceland (Table 1), and will be formally described elsewhere, once its morphological circumscription and geographical distribution are better understood (T. Goward & J. Miadlikowska, unpublished data).

The final taxon collected at Öskjuhlid comprised six representatives of an undescribed species, formally described here as *Peltigera islandica* sp. nov. (see below). Five specimens (SSM55, SSM63, SSM165, SSM348, SSM355) were distinctly emerald green when wet (Fig. 2A–C), but the sixth (SSM344), which shared an identical ITS sequence, was dark green (Fig. 2G) and is interpreted as a variant (colourmorph) of the species. These lichens occurred as small patches of a few lobes directly on soil and were found on both sides of the path, alone or with other species in the same or adjacent sectors (Fig. 1B).

Additional bright green-coloured examples of the new species were collected from three other locations in south-western Iceland (Fig. 1A, Table 1). In each instance, the collecting locality had been subject to recent disturbance: on a residential property in a neighbourhood (Thverás) in Reykjavík (OSA332, OSA392; Fig. 2D–F), beside a major road near Hrauneyjafoss (SW329, SW330) in the south highlands, and on the bank of the River Thjórsá (SW331). The ITS sequences from all these specimens were identical to that of SSM55.

Phylogenetic delimitation of *P. islandica* mycobiont

The ITS has proved to be a very informative molecular marker for species delimitation in the genus *Peltigera* (Goffinet *et al.* 2003;

Miadlikowska *et al.* 2014b; Han *et al.* 2015), and the 8–126 bp hypervariable region in the ITS1 (ITS1–HR) has been especially useful for species distinction in the taxonomically very challenging section *Peltigera* (Miadlikowska *et al.* 2003). As the ITS sequence of accession SSM55 was 99% identical (546/547 nucleotides) to that deposited for *P. sp. A* (GenBank FJ708907, undescribed Canadian material represented by HOB020708–62–1–1; O'Brien *et al.* 2009), a context for the phylogenetic validation of *P. islandica* could be established based on the strong bootstrap support that *P. sp. A* had received for placement in the section *Peltigera*, Group C (*'P. canina* group'; O'Brien *et al.* 2009) (Fig. 3).

Although not all relationships among species in the present phylogeny of Group C are settled, most of the species represented by multiple individuals were significantly supported as monophyletic (Fig. 3): *P. praetextata*, *P. koponenii*, and the putative *P. "fuscopraetextata"* (including *P. "pallidorufescens"*) (Miadlikowska *et al.* 2003; O'Brien *et al.* 2009). Length variation of homopolymeric tracts in the ITS1–HR region within the *P. "fuscopraetextata"/P. "pallidorufescens"* clade may indicate the occurrence of multiple species or a single broadly defined taxon (see O'Brien *et al.* 2009). Similarly, shared C/T substitutions in ITS1–HR between some individuals of *P. canina* (e.g. AY257952) and *P. "boreorufescens"* (e.g. AY257896), and their position outside the *P. canina* s. str./*P. koponenii* clade, may indicate a new species (*P. "boreorufescens"*) or intraspecific variation within a broadly defined *P. canina*, if its monophyly is confirmed (see Miadlikowska *et al.* 2003). The validity of *P. koponenii* (known only from Papua New Guinea; Sérusiaux *et al.* 2009) as an independent species versus a morphotype within *P. canina* s. str., as suggested by its identical ITS1–HR with members of *P. canina* s. str., should also be further evaluated.

Against this background, all individuals of *P. islandica* and *P. sp. A* examined stand together for their lack of the ITS1–HR, which is present in all other members of the

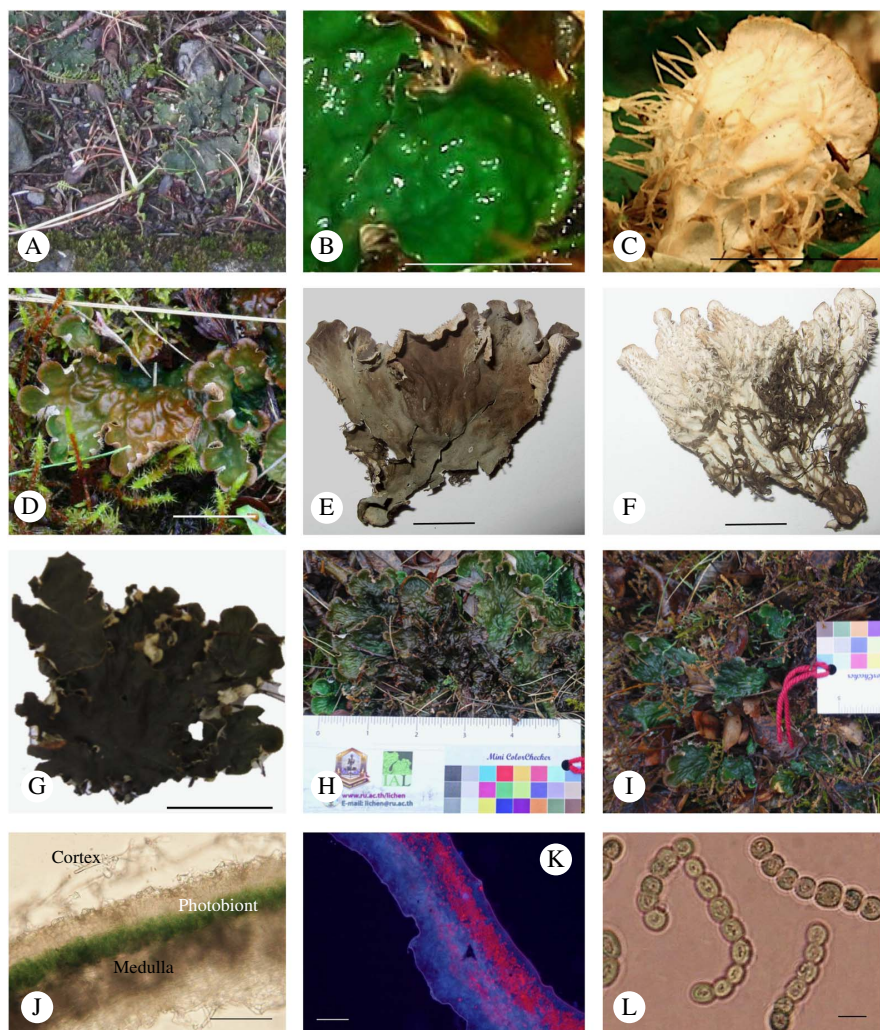


FIG. 2. *Peltigera islandica*. A, voucher SSM348; B & C, voucher SSM63, upper and lower surface, respectively; D–F, voucher OSA392, *in situ* and upper and lower surface of a portion of the dried holotype; G, voucher SSM344; H & I, thallus from which voucher OSA392 was collected (April 2016), and a new patch nearby; J & K, SSM63 thallus cross-sections viewed under white and UV light, respectively; L, *Nostoc* SSM165-N2a. Scales: B–F = 1 cm; G = 2 cm; J & K = 50 μ m; L = 10 μ m. A ruler with colour card (7th International Association of Lichenology Symposium, Bangkok, 2012) was used as a standard to cross reference between images H and I.

P. canina complex. As expected, *P. islandica* and *P. sp.* A form a monophyletic, highly supported separate lineage, although here

a sister relationship with *P. praetextata* (see O'Brien *et al.* 2009) was not recovered. Overall bootstrap support was in agreement

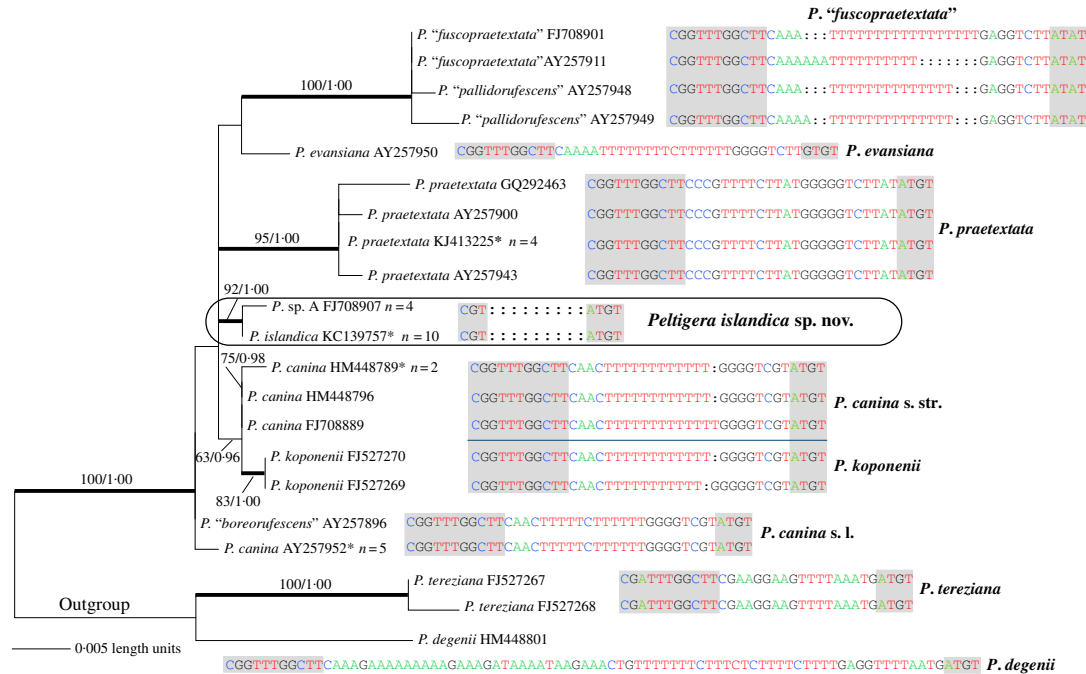


FIG. 3. Phylogenetic relationships among selected members (7 taxa, 18 OTUs) of the *P. canina* complex (section *Peltigera*) based on ITS sequences. Names in quotation marks indicate putative undescribed species. Sequences of the hypervariable ITS1 region (absent from *P. islandica*) are aligned within each putative species; conserved flanking motifs are shaded in grey. Values associated with internodes represent bootstrap support (ML-BP; before the slash) and posterior probability support (PP; after the slash). Only values >50% for ML-BP and 0.95 for PP are shown. Thick branches indicate relationships with strong bootstrap (ML-BP $\geq 70\%$) and posterior probability (PP $\geq 0.95\%$) support. *Peltigera degenii* and *P. tereziana* were used to root the tree. * = includes newly generated sequences from specimens collected in Iceland. In colour online.

with posterior probability support (both or neither significant) for all internodes except for the one in which *P. koponenii* was sister to *P. canina* s. str., with the exclusion of *P. canina* s. lat. (significant support by posterior probability only; Fig. 3). Contrary to other studies (Gaya *et al.* 2011; Sliwa *et al.* 2012), incorporating signal from ambiguously aligned regions in the ITS and using an alignment-free method (BALi-Phy) did not increase the resolution and robustness of the ITS phylogeny, perhaps due to an insufficient number of variable characters provided by the ITS locus, even if ambiguously aligned regions (delimited manually and excluded from the RAxML analyses) were incorporated in the Bayesian analysis.

Two other mycobiont loci were also examined. *RPB1* was the same among all accessions of *P. islandica* (Table 1) and *P. sp. A* (represented by GenBank FJ709247 from HOB020708–62–1–1). All *P. islandica* samples shared the same *tub-2* allele, but this differed by two or three silent mutations from the alleles reported for *P. sp. A* (represented by GenBank FJ709454 and FJ709453 from HOB020708–62–1–4 and HOB020708–62–1–1, respectively). Nonetheless, all three were distinct from alleles of other members of the '*P. canina*' group, including those newly obtained for this study.

The sister relationship of *P. islandica* and *P. sp. A* revealed by the ITS phylogeny (Fig. 3) and identity or close similarity of other genes indicate that the emerald green specimens, as well as the dark green forms most likely including *P. sp. A*, represent a single new species. This new member of the *P. canina* species complex (section *Peltigera*) is described here as *P. islandica*.

Characterization of photobiont

Hand-sectioned water mounts of *P. islandica* SSM55 and SSM63 (Fig. 2J & K) examined by light microscopy revealed the stratified interior expected in bipartite *Peltigera* species. The photobiont layer (150–175 µm) appeared green under white light, but autofluoresced red under UV illumination, consistent with the presence of chlorophylls

and/or phycobilins. *Nostoc*-like cells, 6–10 µm in diameter, were present in short chains of several cells. The 16S rRNA gene sequences amplified from whole lichen DNA, extracted from accessions SSM55 (KC139748) and SSM63, were identical and shared 99% identity with sequences from symbiotic *Nostoc* sp. from plants (e.g. 1271/1274 nucleotide identity to *Nostoc* from *Gunnera tinctoria*, GenBank AM711527; Papaefthimiou *et al.* 2008) and lichens (e.g. 1270/1274 identity to *Nostoc* sp. from *P. extenuata*, GenBank KF359710.1).

Results from the examination of whole lichens were supported by characterization of cyanobacteria cultivated from *P. islandica* SSM165. The isolates, SSM165-N2a and SSM165-N2b, were consistent with *Nostoc* sp. in morphology (Fig. 2L), and their *rbcLX* sequences (KC139777, KC139778), identical to that of the source lichen SSM165 (haplotype *rbcLX*-1, Table 1), shared high similarity to reported *Nostoc* photobionts (below).

The *rbcLX* locus was useful not only for supporting identification of the photobiont to *Nostoc* sp., but also for distinguishing strains and placing the photobiont of *P. islandica* in a larger biological and phylogenetic context. Of the nine *rbcLX* haplotypes associated with the *Peltigera* community at Öskjuhlid (Table 1, Fig. 1), all five emerald green forms of *P. islandica* had *rbcLX*-1, while the dark green form, SSM344, had *rbcLX*-2, a quite different allele with a 24-nucleotide insertion and >10 substitutions compared to *rbcLX*-1. These haplotypes were not exclusive to *P. islandica*: *rbcLX*-1, the most common allele at Öskjuhlid, was present in *P. rufescens* OSA451 as well as four of five *P. "neorufescens"*, while *rbcLX*-2 was also found with *P. rufescens* SSM337 nearby. At Hrauneyjafoss, one *P. islandica* (SW329) shared *rbcLX*-1 with the green morphotypes at Öskjuhlid, but the second (SW330), and those from the final two localities were different, with changes at 2, 4 and 9 positions in *rbcLX*, respectively (Table 1).

Despite the diversity of *rbcLX* types associated with colourmorphs and locations, when compared to a large worldwide dataset

that included a significant proportion of *Peltigera* symbionts (O'Brien *et al.* 2013), all photobionts of *P. islandica* grouped together in a large clade (phylogroup I; Fig. 4) that

also contained the majority of *Peltigera rbcLX* types known from Iceland, including many from *P. membranacea* (Manoharan *et al.* 2012). Within this clade, cyanobionts from

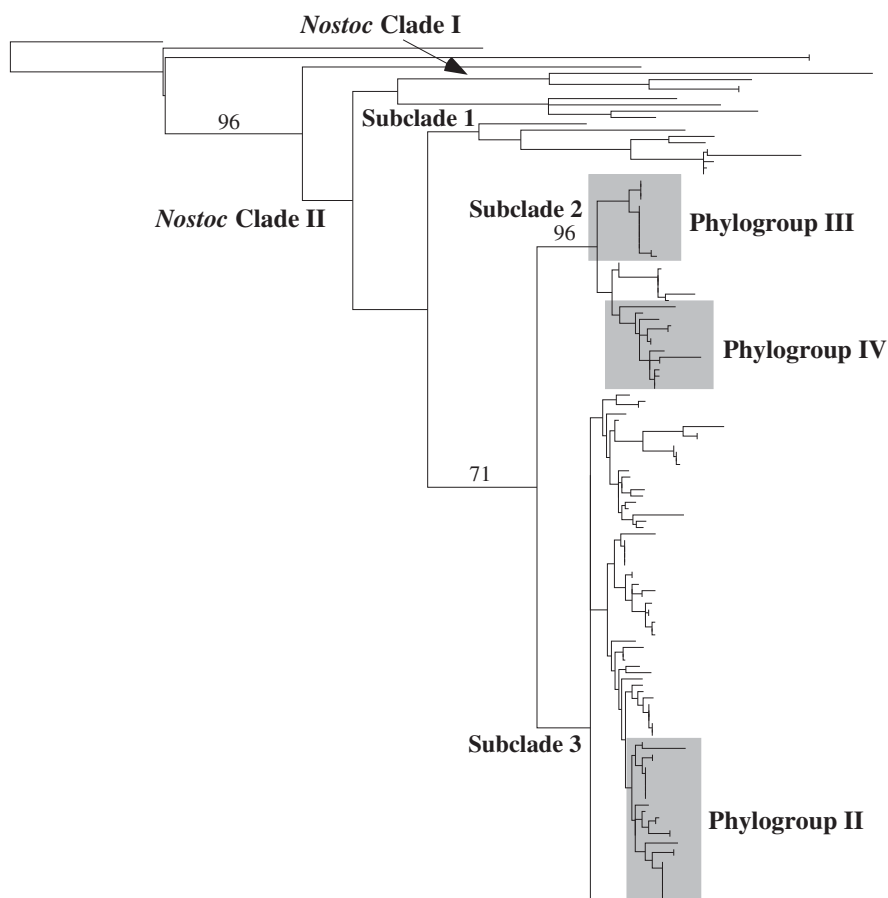


FIG. 4. Phylogenetic placement of cyanobionts associated with *P. islandica* (including *P. sp. A*) and other *Peltigera* species from Iceland revealed by maximum likelihood analysis of 295 individuals based on the *rbcLX* locus. Names of all Icelandic specimens and *P. sp. A* are in bold. Boxes identify phylogroups (I, VI, and V) that include the *Nostoc* symbionts from Iceland (a complete phylogeny is available upon request). Names of Icelandic specimens are appended with a hyphen and corresponding *rbcLX* allele designation. *n* = 1 unless otherwise stated. ● = cyanobionts from dark green thalli of *P. islandica* from Iceland; ■ = cyanobionts from dark green thalli of *P. islandica* (including *P. sp. A*) from Canada; ○ = *P. islandica* cyanobionts from emerald green thalli collected in Iceland. Bootstrap support values $\geq 70\%$ are provided above internodes. Delimitation of clades, subclades, and phylogroups (clusters) follows Otálora *et al.* (2010) and O'Brien *et al.* (2013). *Fischerella muscicola* (*Stigonematales*) was used to root the tree.

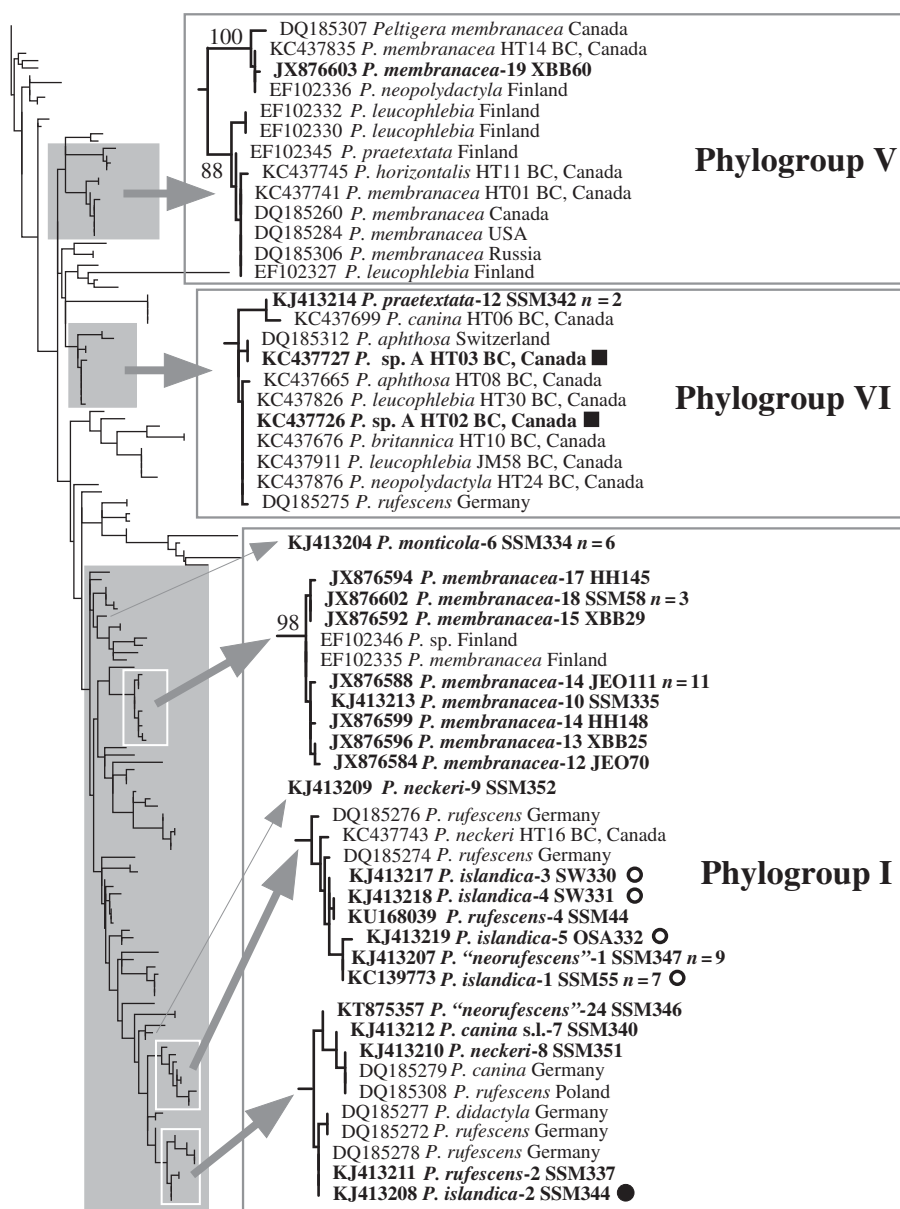


FIG. 4. (continued)

the emerald green morphs of *P. islandica* were clustered, regardless of geographical origin, whereas *Nostoc* from the dark green morph of *P. islandica* and *P. rufescens* SSM337, in a different subclade, was more closely related to strains from geographically distant (e.g. from Germany and Poland) species of section *Peltigera*. In contrast, cyanobionts of *P. membranacea* (with one exception) seem to be associated with a unique (highly supported), potentially widely distributed (across Iceland and beyond, e.g. Finland) clade of *Nostoc*. Notably, the photobionts reported with the Canadian specimens of *P. sp. A* were placed in a different cluster (phylogroup VI; Fig. 4), despite availability, albeit in low abundance (2%), of phylogroup I photobionts at their collection sites in Canada (O'Brien *et al.* 2013).

Taxonomic Section

***Peltigera islandica* Goward & Manoharan-Basil sp. nov.**

Mycobank No.: MB 817007

Thallus foliose, some forms similar to *P. membranacea*, others to *P. rufescens*, but lobes narrower than the first and lobe tips predominately distinctly downturned, not exclusively upturned as in the second. The hypervariable region in the internal transcribed spacer 1 (ITS1-HR; Miadlikowska *et al.* 2003) is absent, contrary to the remaining species currently recognized in the section *Peltigera*.

Type: Iceland, Thverás, terricolous, 64°10'41.83"N, 21°7'80.667"W, with grass and moss, under *Salix arctica* and *Calluna vulgaris* by a driveway in a residential neighbourhood, 5 August 2015, Ó. S. Andrésson OSA392 (AMNH—holotype; UBC, H—isoatypes).

(Fig. 2D–F)

Thallus foliose, loosely appressed, usually small, 1.5–3.0 cm across (see below); *lobes* thin, brittle, 5–10(–15) mm wide, loosely overlapping, short to somewhat elongate, irregularly branching; *lobe tips* rounded, predominantly downturned; *lobe margins* mostly even, not crisped; *upper surface* bright emerald green to apparently rarely khaki green (moist) or pale bluish grey to purplish brown (dry), dull, often distinctly billowed, tomentose near lobe tips, the tomentum

becoming sparse or absent toward thallus centre, appressed to occasionally in part erect; *soredia*, *isidia* and *marginal lobules* lacking; lower surface distinctly veined; *veins* whitish except grading inward to medium brown, not appearing to overlap, narrow to somewhat broad, low to more often raised, usually lacking tomentum, but sometimes distinctly erect-tomentose; *interstices* white, lenticular to occasionally in part polygonal, moderately deeply set; *rhizines* concolourous with veins, variable, ranging from discrete to flocculent or hedgerow-forming towards lobe tips, inwards becoming mostly threadlike and unbranched, 2–3(–5) mm long, lacking tomentum or sometimes distinctly erect-tomentose; *cortex* c. 250–300 µm thick; *photobiont layer* 150–175 µm thick, containing *Nostoc*; *medulla* white, c. 450–550 µm thick.

Apothecia not seen.

Chemistry. All spot tests negative.

Distribution. *Peltigera islandica* is known from Iceland and from southern intermontane British Columbia, Canada.

Discussion. *Peltigera islandica* is characterized by the presence of a cyanobacterial photobiont (*Nostoc*) in combination with a usually bright emerald green upper surface when wet, narrow lobes 5–10 mm wide, laminal tomentum, and predominantly downturned lobe tips. *Peltigera membranacea*, *P. praetextata* and to some extent *P. rufescens*, are similar but the first species differs in having broader lobes (10–20(–30) mm wide) and the latter two species differ in having upturned lobe tips.

This is perhaps one of the most variable members of section *Peltigera* (the '*P. canina* group'), especially with respect to the morphology of the lower surface. In some thalli, for example, the veins and rhizines are densely covered in erect tomentum, whereas tomentum is lacking altogether in other thalli, with every gradation between. The outermost rhizines too can be hedgerow-forming (as in *P. rufescens*), to threadlike/non-tomentose (as in *P. praetextata*), flocculent (as in *P. canina*), or threadlike/erect-tomentose

(as in *P. membranacea*), even in different portions of a single thallus. Even so, these species are unlikely to be confused with *P. islandica* owing to the presence of upturned lobe tips in *P. rufescens* and *P. praetextata*, and broader lobes in *P. canina* and *P. membranacea*. In contrast to *P. islandica*, all of these currently recognized species in section *Peltigera* contain ITS1-HR, as defined in Miadlikowska *et al.* (2003).

The material of *P. islandica* from which lobes were sampled for accessions OSA332 and OSA392 (the holotype) was found on previously bare soil beside a residential driveway built in 1989. Those vouchers are from a larger colony than those at Öskjuhlid, but apparently still represent one thallus that was first noted in 2013 for its emerald green colour. The same thallus was re-photographed *in situ* most recently in April 2016 (Fig. 2H), and estimated to be c. 12 cm across. The oldest (central) part has deteriorated, with lobes that are broken and brown; growth at the periphery, however, is emerald green. No apothecia were noted, but another smaller, emerald green patch was found nearby (Fig. 2I). We made the decision not to collect all of the original patch for use as holotype material in order to maintain the type strain of the mycobiont (and possibly the photobiont) in nature, leaving open the possibility of obtaining additional clonally-related material for archival as well as for future studies.

Selected specimens examined. Iceland: Öskjuhlid: 64°13'16.83"N, 21°9'20.17"W, terricolous with *Hylocomium splendens* and other mosses, beside a paved path in an afforested area with *Betula pubescens*, *Picea sitchensis* and *Pinus mugo*, 2013, S. S. Manoharan-Basil SSM344, SSM348 and SSM355 (IMNH). *Thverás:* 64°10'41.83"N, 21°78'06.67"W, terricolous with grass and moss, under *Salix arctica* and *Calluna vulgaris*, by a driveway in a residential neighbourhood, 2013, Ö. S. Andrússon OSA332 (IMNH).—*Canada:* British Columbia: Spahats Falls, 2400 m, 51°7'38.889"N–120°0'16.111"W, on soil above trail along W-facing slope, N of waterfall in *Abies-Betula* forest, with *Thuja* and *Rosa* understorey, 2002, P. sp. A H. O'Brien HOB020708-66-1-4 (DUKE).

Discussion

The lichen flora of Iceland has been surveyed since 1870 and was noted to comprise

over 755 species by 2009 (Kristinsson & Heiðmarsson 2009). While it is considered fairly well studied by lichenologists, new species continue to be added. Hansen (2009) conducted floristic studies of nine localities in 2007 and 2008 and documented 12 species new to Iceland. Kristinsson *et al.* (2014) reviewed the herbarium of Svanhildur Svane and added two more, on top of the 15 from this collection already provisionally incorporated (Kristinsson & Heiðmarsson 2009). The present study adds a further two new entries to the list of Icelandic lichens, and brings the number of *Peltigera* species recorded there to 24.

Peltigera species worldwide are often considered colonizers of areas with man-made or natural disturbance (Topham 1977; Louwhoff 2009), and this seems generally consistent with where the new taxa in this study were found. The locations in south-west Iceland are geographically separated so dispersal has occurred, although the mechanism(s) are presently unknown. Clonal reproduction (vertical transmission of the photobiont) as a means to account for the presence of *rbcLX*-1 throughout the population of *P. islandica* with emerald green thalli at Öskjuhlid is an attractive hypothesis for its simplicity, although formally, high photobiont selectivity for the species cannot be ruled out (Otalora *et al.* 2010; O'Brien *et al.* 2013). The latter may be unlikely, however, as the mycobiont is also associated with *Nostoc rbcLX*-2 at Öskjuhlid, as well as other *rbcLX* types at each of the other locations (whether by *de novo* synthesis or "photobiont sharing" (Rikinen *et al.* 2002; Rikinen 2003) is unclear). Low selectivity may characterize this species more generally, as the specimens originally designated *P. sp. A* from Canada carry *Nostoc* strains from a different phylogroup to those in Iceland (O'Brien *et al.* 2013).

It may be more relevant to note that in each case the association appeared to be with a dominant local photobiont: at Öskjuhlid *rbcLX*-1 was common, while in Canada the haplotype that included the *P. sp. A* photobionts comprised about two-thirds of all *Nostoc* determined from the sites

(O'Brien *et al.* 2013). The capacity to adapt to locally abundant and presumably successful photobionts could confer an advantage for exploiting a variety of environments, perhaps facilitating expansion of the geographical range of the mycobionts. In connection with this, a 99% match to an ITS sequence (GenBank accession no. KC965216) from a metagenomic study suggests that *P. islandica* may also exist in the Canadian Arctic (Timling *et al.* 2014). Overall, occurrence in localities separated by >100 km and association with different local cyanobionts suggests that *P. islandica* is well established in Iceland; however, its status in climatically similar regions of western North America needs further work, and it should also be looked for in Europe.

Characterization of a new species from the onset with consideration of its community as a broader framework has a number of benefits, because a more comprehensive dataset might lead to additional insights into the biology and ecology of the lichen (Seaward 1977; Nash 2008). With adequate sampling, co-occurrence of closely-related species could offer opportunities to explore species boundaries by searching for evidence of gene flow, as was done for mixed communities in British Columbia, wherein *P. sp. A* and *P. canina* were first suggested to be distinct (O'Brien *et al.* 2009). Given that *P. canina* s. lat. from Vaglaskógur and Hrafnhólar characterized in this study have the same *RBP1* and *tub-2* alleles as *P. canina* in British Columbia, both of which are different from those of *P. islandica*, further examination of *P. canina* at Öskjuhlid would allow comparable analyses to be made not only within site in Iceland, but also across continents.

From an ecological perspective, the presence of multiple *Peltigera* species in the study area at Öskjuhlid might represent an initial stage of colonization by ruderals (Topham 1977; Scheidegger 1995; Clayden 2010), in which *P. islandica* would seem at least as adept as the others. Alternatively, it may be interpreted as indicating that *P. islandica* can be an efficient competitor in later stages of ecological succession, insofar

as one study has shown that mature forests hold a higher diversity of *Peltigera* species than young forests (Ramirez-Fernandez *et al.* 2013). In Iceland, the original forests were felled centuries ago and nearly all large tracts with trees today result from afforestation since the mid-1900s (Eysteinnsson 2013). Relative to stand age, the concrete path at Öskjuhlid is therefore not a recent disturbance, but as much a fixture as the trees, and the 'snapshot' of the consortia of *Peltigera* beside it might represent a seral stage of transition. Further characterization and monitoring of this and other terrestrial lichen communities in modern Icelandic forests might offer an additional window into this emerging landscape.

A second important realization is that *P. islandica* is probably more common in Iceland than implied by its recent recognition as a new species. Although colour is not typically a reliable character for bipartite *Peltigera* species, the distinctive emerald green hue, coupled with the absence of cephalodia, was nonetheless part of the original search criteria, and not surprisingly only green morphs were recovered from the other locations. The dark green morph would have been overlooked, except in studies where the goal was to examine the community. While this form is a minority at Öskjuhlid, a more thorough survey is needed to accurately determine the abundance and distribution of *P. islandica* in all its forms in Iceland. The prevalence of the green form might be related to a humid microhabitat and/or association with particular cyanobionts (e.g. *Nostoc rbcLX-1*, *rbcLX-6*), as there were a number of other somewhat greenish bipartite *Peltigera* species at Öskjuhlid. Among them, the finding of *P. "neorufescens"*, another taxon new to Iceland, was interesting, given that some thalli were located within a metre of *P. islandica*. The detection of two ITS haplotypes of *P. "neorufescens"* in Iceland suggests that it also might be relatively common.

The number and the circumscriptions of *Peltigera* species worldwide are likely to change as a result of recent lichen inventories

(e.g. China and South America) and ongoing revisionary studies with extended sampling and new *Peltigera*-specific markers. With the use of a more integrated approach to complement traditional morphology-based investigations of new species, exploration of even small spatial scale *Peltigera* communities in interesting habitats might also offer opportunities to discover and study yet more additions to this genus.

We thank S. Werth for the gift of samples SW329–331, B. Redeling for help in performing BALI-Phy analysis, and N. Magain and C. Truong for providing a broad phylogenetic context (unpublished data) for the validation of *P. islandica*. We are also very grateful to our reviewers from the journal for their useful suggestions. JM was supported by the National Science Foundation through a REVSYS award on the genus *Peltigera* (DEB-1025930); SSM and ÖSA were supported by the Icelandic Research Fund.

The authors dedicate this paper to Dr Irwin M. Brodo, in recognition of his many important contributions to lichen taxonomy and biology over the past half century.

REFERENCES

- Clayden, S. R. (2010) Lichens and allied fungi of the Atlantic Maritime Ecozone. In *Assessment of Species Diversity in the Atlantic Maritime Ecozone* (D. F. McAlpine & I. M. Smith, eds): 153–178. Ottawa: NRC Research Press.
- Eysteinnsson, T. (2013) *Forestry in a Treeless Land*. Egilsstaðir, Iceland: Iceland Forest Service.
- Gaya, E., Redelings, B. D., Navarro-Rosines, P., Llimona, X., De Cáceres, M. & Lutzoni, F. (2011) Align or not to align? Resolving species complexes within the *Caloplaca saxicola* group as a case study. *Mycologia* **103**: 361–378.
- Glass, N. L. & Donaldson, G. C. (1995) Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Applied and Environmental Microbiology* **61**: 1323–1330.
- Goffinet, B. & Bayer, R. J. (1997) Characterization of mycobionts of photomorph pairs in the Peltigerineae (lichenized ascomycetes) based on internal transcribed spacer sequences of the nuclear ribosomal DNA. *Fungal Genetics and Biology* **21**: 228–237.
- Goffinet, B., Miadlikowska, J. & Goward, T. (2003) Phylogenetic inferences based on nrDNA sequences support five morphospecies within the *Peltigera didactyla* complex (lichenized Ascomycota). *Bryologist* **106**: 349–364.
- Goward, T., Goffinet, B. & Vitikainen, O. (1995) Synopsis of the genus *Peltigera* (lichenized Ascomycetes) in British Columbia, with a key to the North American species. *Canadian Journal of Botany* **73**: 91–111.
- Hainfried, E. & Schenk, A. (2013) Cyanobacterial symbioses. In *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications. 2nd Edition* (A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K. Schleifer, eds): 3819–3854. New York: Springer-Verlag.
- Han, L.-F., Zhang, Y.-Y. & Guo, S.-Y. (2013) *Peltigera wulingensis*, a new lichen (Ascomycota) from north China. *Lichenologist* **45**: 329–336.
- Han, L.-F., Zheng, T.-X. & Guo, S.-Y. (2015) A new species in the lichen genus *Peltigera* from northern China based on morphology and DNA sequence data. *Bryologist* **118**: 46–53.
- Hansen, E. S. (2009) A contribution to the lichen flora of Iceland. *Folia Cryptogamica Estonica* **46**: 45–54.
- Hasegawa, M., Kishino, H. & Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.
- Kirk, P. M., Cannon, P. F., Minter, D. W. & Stalpers, J. A. (2008) *Dictionary of the Fungi*. Wallingford, UK: CABI.
- Kristinsson, H. & Heiðmarsson, S. (2009) Checklist of lichens in Iceland. <http://www.floraislands.is/PDF-skjol/Flettulisti-2009.pdf>.
- Kristinsson, H., Heiðmarsson, S. & Hansen, E. S. (2014) Lichens from Iceland in the collection of Svanhildur Svane [Islandijos kerpės Svanhildur Svane kolekcijoje]. *Botanica Lithuanica* **20**: 14–18.
- Li, L. A. & Tabita, R. (1997) Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *Journal of Bacteriology* **179**: 3793–3796.
- Louwhoff, S. H. J. J. (2009) *Peltigera*. In *Flora of Australia Vol. 57 Lichens 5*. Victoria: CSIRO Publishing/Australian Biological Resources Study (ABRS). <https://www.anbg.gov.au/abrs/lichenlist/Peltigera%20gen.pdf>.
- Lutzoni, F., Wagner, P., Reeb, V. & Zoller, S. (2000) Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Systematic Biology* **49**: 628–651.
- Maddison, D. R. & Maddison, W. P. (2003) *MacClade: Analysis of Phylogeny and Character Evolution, version 4.07*. Sunderland, Massachusetts: Sinauer Associates.
- Manoharan, S. S., Miao, V. P. W. & Andrésson, O. S. (2012) LEC-2, a highly variable lectin in the lichen. *Symbiosis* **58**: 91–98.
- Martinez, I., Burgaz, A. R., Vitikainen, O. & Escudero, A. (2003) Distribution patterns in the genus *Peltigera* Willd. *Lichenologist* **35**: 301–323.
- Matheny, P. B., Liu, Y. J., Ammirati, J. F. & Hall, B. D. (2002) Using RPB1 sequences to improve phylogenetic inference among mushrooms (*Inocybe*, *Agaricales*). *American Journal of Botany* **89**: 688–698.
- Miadlikowska, J. & Lutzoni, F. (2000) Phylogenetic revision of the genus *Peltigera* (lichen-forming Ascomycota) based on morphological, chemical, and large subunit nuclear ribosomal DNA data. *International Journal of Plant Sciences* **161**: 925–958.
- Miadlikowska, J. & Lutzoni, F. (2004) Phylogenetic classification of peltigerean fungi (*Peltigerales*,

- Ascomycota) based on ribosomal RNA small and large subunits. *American Journal of Botany* **91**: 449–464.
- Miadlikowska, J., Lutzoni, F., Goward, T., Zoller, S. & Posada, D. (2003) New approach to an old problem: incorporating signal from gap-rich regions of ITS and rDNA large subunit into phylogenetic analyses to resolve the *Peltigera canina* species complex. *Mycologia* **95**: 1181–1203.
- Miadlikowska, J., Richardson, D., Magain, N., Ball, B., Anderson, F., Cameron, R., Lendemer, J., Truong, C. & Lutzoni, F. (2014a) Phylogenetic placement, species delimitation, and cyanobiont identity of endangered aquatic *Peltigera* species (lichen-forming Ascomycota, Lecanoromycetes). *American Journal of Botany* **101**: 1141–1156.
- Miadlikowska, J., Kauff, F., Hognabba, F., Oliver, J. C., Molnar, K., Fraker, E., Gaya, E., Hafellner, J., Hofstetter, V., Gueidan, C. *et al.* (2014b) A multi-gene phylogenetic synthesis for the class Lecanoromycetes (Ascomycota): 1307 fungi representing 1139 infrageneric taxa, 317 genera and 66 families. *Molecular Phylogenetics and Evolution* **79**: 132–168.
- Miao, V. P. W., Rabenau, A. & Lee, A. (1997) Cultural and molecular characterization of photobionts of *Peltigera membranacea*. *Lichenologist* **29**: 571–586.
- Miller, M. A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, Louisiana*, pp. 1–8.
- Monacelli, J. T. & Carbone, I. (2014) Moby SNAPP: a web-based analysis portal for population genetics and evolutionary genomics. *Bioinformatics* **30**: 1488–1490.
- Nash, T. H., III (2008) *Lichen Biology, 2nd Edition*. Cambridge: Cambridge University Press.
- Nylander, J. A. A. (2004) *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- O'Brien, H. E., Miadlikowska, J. & Lutzoni, F. (2009) Assessing reproductive isolation in highly diverse communities of the lichen-forming fungal genus *Peltigera*. *Evolution* **63**: 2076–2086.
- O'Brien, H. E., Miadlikowska, J. & Lutzoni, F. (2013) Assessing population structure and host specialization in lichenized cyanobacteria. *New Phytologist* **198**: 557–566.
- Orange, A., James, P. W. & White, F. J. (2001) *Microchemical Methods for the Identification of Lichens*. London: British Lichen Society.
- Otalora, M. A., Martinez, I., O'Brien, H., Molina, M. C., Aragon, G. & Lutzoni, F. (2010) Multiple origins of high reciprocal symbiotic specificity at an intercontinental spatial scale among gelatinous lichens (Collemataceae, Lecanoromycetes). *Molecular Phylogenetics and Evolution* **56**: 1089–1095.
- Papaefthimiou, D., Hrouzek, P., Mugnai, M. A., Lukesova, A., Turicchia, S., Rasmussen, U. & Ventura, S. (2008) Differential patterns of evolution and distribution of the symbiotic behaviour in nostocacean cyanobacteria. *International Journal of Systematic and Evolutionary Microbiology* **58**: 553–564.
- Ramirez-Fernandez, L., Zuniga, C., Mendez, M. A., Caru, M. & Orlando, J. (2013) Genetic diversity of terricolous *Peltigera* cyanolichen communities in different conservation states of native forest from southern Chile. *International Microbiology* **16**: 243–252.
- Redelings, B. D. & Suchard, M. A. (2005) Joint Bayesian estimation of alignment and phylogeny. *Systematic Biology* **54**: 401–418.
- Redelings, B. D. & Suchard, M. A. (2007) Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evolutionary Biology* **7**: 40.
- Rikkinen, J. (2003) Ecological and evolutionary role of photobiont-mediated guilds in lichens. *Symbiosis* **34**: 99–110.
- Rikkinen, J., Oksanen, I. & Lohtander, K. (2002) Lichen guilds share related cyanobacterial symbionts. *Science* **297**: 357.
- Rodriguez, F., Oliver, J. L., Marin, A. & Medina, J. R. (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485–501.
- Rudi, K., Skulberg, O. M. & Jakobsen, K. S. (1998) Evolution of cyanobacteria by exchange of genetic material among phylogenetically related strains. *Journal of Bacteriology* **180**: 3453–3461.
- Seaward, M. R. D. (ed.) (1977) *Lichen Ecology*. London: Academic Press.
- Scheidegger, C. (1995) Reproductive strategies in *Vezdaea* (Lecanorales, lichenized Ascomycetes): a low-temperature scanning electron microscopy study of a ruderal species. *Cryptogamic Botany* **5**: 163–171.
- Sérusiaux, E., Goffinet, B., Miadlikowska, J. & Vitikainen, O. (2009) Taxonomy, phylogeny and biogeography of the lichen genus *Peltigera* in Papua New Guinea. *Fungal Diversity* **38**: 185–224.
- Sigurðsson, H. M. L. (1993) *Öskjuhlíð, Náttúra og Saga*. Reykjavík, Iceland: Árbæjarsafn og Borgarskipulag.
- Sinnemann, S. J., Andresson, O. S., Brown, D. W. & Miao, V. P. W. (2000) Cloning and heterologous expression of *Solorina crocea* pyrG. *Current Genetics* **37**: 333–338.
- Sliwa, L., Miadlikowska, J., Redelings, B. D., Molnar, K. & Lutzoni, F. (2012) Are widespread morphospecies from the *Lecanora dispersa* group (lichen-forming Ascomycota) monophyletic? *Bryologist* **115**: 265–277.
- Stamatakis, A. (2006) RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* **57**: 758–771.
- Stiller, J. W. & Hall, B. D. (1997) The origin of red algae: implications for plastid evolution. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 4520–4525.
- Suchard, M. A. & Redelings, B. D. (2006) BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**: 2047–2048.
- Timling, I., Walker, D. A., Nusbaum, C., Lennon, N. J. & Taylor, D. L. (2014) Rich and cold: diversity,

- distribution and drivers of fungal communities in patterned-ground ecosystems of the North American Arctic. *Molecular Ecology* **23**: 3258–3272.
- Topham, P. B. (1977) Colonization, growth, succession and competition. In *Lichen Ecology* (M. R. D. Seaward, ed.): 31–68. London: Academic Press.
- Turner, S., Pryer, K. M., Miao, V. P. W. & Palmer, J. D. (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology* **46**: 327–338.
- Vitikainen, O. (1994) Taxonomic revision of *Peltigera* (lichenized Ascomycotina) in Europe. *Acta Botanica Fennica* **152**: 1–96.
- Vitikainen, O. (2006) *Peltigera tartarea*, a new species from Arctic America. *Journal of the Hattori Botanical Laboratory* **100**: 853–854.
- White, T., Bruns, T., Lee, S. & Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White, eds): 315–322. San Diego: Academic Press.
- Wilgenbusch, J. C., Warren, D. L. & Swofford, D. L. (2004) AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty>.

4 SUMMARY AND OUTLOOK

Characterization of the macromolecular components of a lichen can significantly promote the understanding of lichen physiology and evolution, and identify genes and proteins involved in symbiotic interactions and specialization relating to the lichenized habitat. Lichens can be recalcitrant to culture, but the uncultivated genome can be studied by sequencing a field sample. Subsequently characterizing the tissue-specific (symbiotic and aposymbiotic) transcriptome and methylome components provides unique opportunities and form the basis to identify genes that are differentially expressed, provides insight into nutrient cycling, gene content and functions, metabolic networks, and the whole genome methylome. *Peltigera membranacea*, the model organism used in this study, is a terricolous lichen, has a symbiotic association with *Nostoc* sp. cyanobacteria and is recalcitrant to culture. These lichens are distributed worldwide and are important contributors to the ecosystems, being involved in nutrient assimilation (carbon and nitrogen cycling) and providing habitat for other organisms. In this dissertation, the tissue-specific transcriptome and methylome of this lichen were studied. It also involves genome annotation and characterization of a new cyanolichen species.

For the transcriptome study (I), the first task was to obtain good quality RNA that can be used for high-throughput sequencing. Several RNA extraction methods were tested to yield high-quality RNA, and RNA was extracted from the symbiotic (thallus), aposymbiotic vegetative (rhizines) and aposymbiotic reproductive (apothecia) tissues. The draft genome of *P. membranacea* was integrated with the transcriptome data and annotation was validated. *De novo* transcriptomes were generated from the symbiotic and aposymbiotic tissues, and the transcripts were used as ESTs in the automatic annotation of the genome. A total of 16,400 genes were annotated. Functional annotation was carried out for the predicted genes using BlastKOALA, GhostKOALA, and Blast2GO. Information on InterProScan (protein domains), EC, and KEGG were added to the functional description. Over one hundred KEGG pathways were predicted in *P. membranacea*, including all pathways of central fungal metabolism with 34 complete pathways.

Reference-based RNA–Seq mapping was carried out for each of the tissues to identify tissue-specific differentially expressed genes (DEGs). Differential gene expression studies were based on tissue samples collected during the years 2009, 2010, 2011 and 2012. A total of nine tissue comparisons were made with a total of 2653 DEGs were identified from the various tissue comparisons. After removal of genes with high background variation in biological replicates, a total of 1552 DEGs remained. Out of 1552 DEGs, 298 genes were without any blast hit, 546, 40 and 17 genes were hypothetical, uncharacterised and predicted proteins. Although, the above genes (~ 901 genes) did not have any functional information they were expressed in the three tissues and thus some of the genes could be unique to lichens and requires further validation. The differentially expressed genes with functional association involving nutrient assimilation, melanin biosynthesis, and the glyoxylate pathway in different tissue types were examined in detail. Potential model of exchange of carbon and nitrogen flow between symbiotic partners is proposed. Genes encoding enzymes for melanin biosynthesis that provide defense against environmental stresses such as UV, radiation, oxidizing agents and ionizing radiation were identified. The DHN–melanin pathway, the DOPA–melanin pathway and the L–tyrosine degradation pathways are all represented in *P. membranacea*. Based on the identification of the two mating type genes (*MAT1-1-1* and *MAT1-2-1*) in a single thallus as well as heterokaryon incompatibility related genes it is surmised that *P. membranacea* is homothallic and also has a parasexual cycle.

Several genes implicated in symbiosis such as lectin genes and hydrophobin genes were identified. Two galectin genes, *lec-1* and *lec-2*, were further characterized (III and IV). Lectins have been thought to be involved in the recognition of photobiont partners in lichen symbioses. The galectin-like genes, *lec-1* and *lec-2* were differentially expressed, *lec-1* was highly expressed in thalli, and *lec-2* was highly expressed in apothecia. *lec-1* is thought to be a determinant in the establishment or maintenance of compatible symbionts whereas *lec-2* was found to be highly polymorphic and under positive selection. Nevertheless, the positive selection of LEC-2 variants did not correlate with variation in photobiont partners. Apart from the above two characterized lectin genes (III and IV), genes encoding additional five fucose-specific lectins, two concanavalin A-like lectins, and a legume like lectin were identified. Also, some genes potentially relevant to symbiosis such as hydantoin

racemase and mitogen-activated kinase *spm1* were differentially expressed. The role of these genes in symbiosis needs to be elucidated.

For the methylome study (II), DNA was extracted from thallus and rhizine tissues. The DNA was treated with bisulfite, and high throughput sequencing was performed. The BS-Seq reads were mapped onto the reference genome and analyzed. Based on the BS-Seq reads generated from thallus whole, thallus body (depleted of rhizines) and rhizines, the overall fraction of methylated Cs in *P. membranacea* was estimated to be ~ 15%. The DNA was methylated in all sequence contexts, i.e., CG, CHG, and CHH. Levels of cytosine methylation at CG, CHG, and CHH nucleotides are about 7.9%, 12.8%, and 16.1%, respectively. Differentially methylated regions in exons and introns were identified. A homolog of *dim-2*, the putative methyltransferase, as well as other genes involved in DNA methylation and RNA-mediated silencing were identified, but the *rid* gene, associated with repeat induced mutations (RIP, found in most ascomycetes hitherto studied), was absent in *P. membranacea* and there was no evidence for RIP (repeat-induced point mutations) from base ratio skewing in REs (repeat elements).

During field sampling, a bright green lichen was collected out of interest and upon further investigation it was characterized as a new lichen species, *Peltigera islandica* (V) in the section *Peltigera*. *P. membranacea* is the first cyanolichen in which the transcriptome and methylome profile of different tissues have been studied. The genome annotation effort combined with the assessment of differential gene expression in the tissues has facilitated the identification of several genes which appear to show tissue-specific expression in the lichen. The findings presented here can be used to formulate new hypotheses and in the planning of new experiments forming a basis for future studies of gene expression, methylation, and function in *P. membranacea*. Also, future directions towards genome finishing (gap closure), and assigning of the scaffolds to chromosomes can be carried out either by additional sequencing or using long read technologies. The global macromolecular analysis presented here can be used as a reference to study other cyanolichens and most importantly the results provided here will help accelerate lichen omics research.

5 REFERENCES

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology* 215 (3):403-410. doi:10.1016/S0022-2836(05)80360-2
2. Antequera F, Tamame M, Villanueva JR, Santos T (1984) DNA methylation in the fungi. *Journal of Biological Chemistry* 259 (13):8033-8036
3. Arber W (1974) DNA modification and restriction. *Progress in nucleic acid research and molecular biology* 14 (0):1-37
4. Armaleo D, Miao V (1999) Symbiosis and DNA methylation in the *Cladonia* lichen fungus. *Symbiosis* 26 (2):143-163
5. Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, Hofstetter V, Kauff F, Lutzoni F (2009) A phylogenetic estimation of trophic transition networks for ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? *Systematic biology* 58 (3):283-297. doi:10.1093/sysbio/syp001
6. Balogh J, Tunlid A, Rosen S (2003) Deletion of a lectin gene does not affect the phenotype of the nematode-trapping fungus *Arthrobotrys oligospora*. *Fungal genetics and biology : FG & B* 39 (2):128-135
7. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. *The Plant journal : for cell and molecular biology* 51 (5):910-918. doi:10.1111/j.1365-3113X.2007.03193.x
8. Barry C, Faugeron G, Rossignol JL (1993) Methylation induced premeiotically in *Ascobolus*: coextension with DNA repeat lengths and effect on transcript elongation. *Proceedings of the National Academy of Sciences of the United States of America* 90 (10):4557-4561
9. Bates ST, Cropsey GW, Caporaso JG, Knight R, Fierer N (2011) Bacterial communities associated with the lichen symbiosis. *Appl Environ Microbiol* 77 (4):1309-1314. doi:10.1128/AEM.02257-10
10. Beckett RP, Minibayeva FV, Liers C (2012) Occurrence of high tyrosinase activity in the non-Peltigeralean lichen *Dermatocarpon miniatum* (L.) W. Mann. *The Lichenologist* 44 (06):827-832
11. Bell-Pedersen D, Dunlap JC, Loros JJ (1992) *The Neurospora*

- circadian clock-controlled gene, *cgc-2*, is allelic to *eas* and encodes a fungal hydrophobin required for formation of the conidial rodlet layer. *Genes & development* 6 (12A):2382-2394
12. Bird A (2007) Perceptions of epigenetics. *Nature* 447 (7143):396-398. doi:10.1038/nature05913
 13. Bomar L, Maltz M, Colston S, Graf J (2011) Directed culturing of microorganisms using metatranscriptomics. *mBio* 2 (2):e00012-00011. doi:10.1128/mBio.00012-11
 14. Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen D, Paietta J, Plesofsky N, Plamann M, Goodrich-Tanrikulu M, Schulte U, Mannhaupt G, Nargang FE, Radford A, Selitrennikoff C, Galagan JE, Dunlap JC, Loros JJ, Catcheside D, Inoue H, Aramayo R, Polymenis M, Selker EU, Sachs MS, Marzluf GA, Paulsen I, Davis R, Ebbole DJ, Zelter A, Kalkman ER, O'Rourke R, Bowring F, Yeadon J, Ishii C, Suzuki K, Sakai W, Pratt R (2004) Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiology and molecular biology reviews* : MMBR 68 (1):1-108
 15. Botton B, Chalot M (1999) Nitrogen Assimilation: Enzymology in Ectomycorrhizas. In: Varma A, Hock B (eds) *Mycorrhiza: Structure, Function, Molecular Biology and Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 333-372. doi:10.1007/978-3-662-03779-9_14
 16. Boyer H (1964) Genetic Control of Restriction and Modification in *Escherichia Coli*. *Journal of bacteriology* 88:1652-1660
 17. Braumann I, van den Berg M, Kempken F (2008) Repeat induced point mutation in two asexual fungi, *Aspergillus niger* and *Penicillium chrysogenum*. *Current genetics* 53 (5):287-297. doi:10.1007/s00294-008-0185-y
 18. Brent MR (2008) Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nat Rev Genet* 9 (1):62-73. doi:10.1038/nrg2220
 19. Brien P, Pugazhendhi D, Woodhouse S, Oxley D, Pell JM (2013) p38alpha MAPK regulates adult muscle stem cell fate by restricting progenitor proliferation during postnatal growth and repair. *Stem cells* 31 (8):1597-1610. doi:10.1002/stem.1399
 20. Bubrick P, Galun M, Frensdorff A (1981) Proteins from the lichen *Xanthoria parietina* which bind to phycobiont cell walls. Localization in the intact lichen and cultured mycobiont. *Protoplasma*:207-211

21. Büdel B, Scheidegger C (2008) Thallus morphology and anatomy. In: Nash III TH (ed) *Lichen Biology*. 2 edn. Cambridge University Press, Cambridge, pp 40-68. doi:DOI: 10.1017/CBO9780511790478.005
22. Bugos RC, Chiang VL, Zhang XH, Campbell ER, Podila GK, Campbell WH (1995) RNA isolation from plant tissues recalcitrant to extraction in guanidine. *Biotechniques* 19 (5):734-737
23. Calmes B, Guillemette T, Teyssier L, Siegler B, Pigne S, Landreau A, Iacomini B, Lemoine R, Richomme P, Simoneau P (2013) Role of mannitol metabolism in the pathogenicity of the necrotrophic fungus *Alternaria brassicicola*. *Frontiers in plant science* 4:131. doi:10.3389/fpls.2013.00131
24. Cambareri EB, Jensen BC, Schabta E, Selker EU (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* 244 (4912):1571-1575
25. Cardinale M, Puglia AM, Grube M (2006) Molecular analysis of lichen-associated bacterial communities. *FEMS Microbiol Ecol* 57 (3):484-495. doi:10.1111/j.1574-6941.2006.00133.x
26. Cardinale M, Vieira de Castro J, Jr., Muller H, Berg G, Grube M (2008) In situ analysis of the bacterial community associated with the reindeer lichen *Cladonia arbuscula* reveals predominance of Alphaproteobacteria. *FEMS Microbiol Ecol* 66 (1):63-71. doi:10.1111/j.1574-6941.2008.00546.x
27. Catalanotto C, Azzalin G, Macino G, Cogoni C (2000) Transcription - Gene silencing in worms and fungi. *Nature* 404 (6775):245-245
28. Catalanotto C, Azzalin G, Macino G, Cogoni C (2002) Involvement of small RNAs and role of the qde genes in the gene silencing pathway in *Neurospora*. *Genes & development* 16 (7):790-795. doi:10.1101/gad.222402
29. Catalanotto C, Pallotta M, ReFalo P, Sachs MS, Vayssie L, Macino G, Cogoni C (2004) Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in *Neurospora crassa*. *Molecular and cellular biology* 24 (6):2536-2545
30. Cernava T, Erlacher A, Aschenbrenner IA, Krug L, Lassek C, Riedel K, Grube M, Berg G (2017) Deciphering functional diversification within the lichen microbiota by meta-omics. *Microbiome* 5 (1):82. doi:10.1186/s40168-017-0303-5
31. Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G,

- Simison M, Miyasato SR, Binkley J, Orvis J, Shah P, Wymore F, Sherlock G, Wortman JR (2014) The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic acids research* 42 (D1):D705-D710. doi:10.1093/nar/gkt1029
32. Cerutti H, Casas-Mollano JA (2006) On the origin and functions of RNA-mediated silencing: from protists to man. *Current genetics* 50 (2):81-99. doi:10.1007/s00294-006-0078-x
33. Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci* 25 (10):481-482
34. Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 6 (5):351-360. doi:10.1038/nrg1601
35. Chang SS, Zhang Z, Liu Y (2012) RNA interference pathways in fungi: mechanisms and functions. *Annu Rev Microbiol* 66:305-323. doi:10.1146/annurev-micro-092611-150138
36. Chapple C (1998) Molecular-Genetic Analysis of Plant Cytochrome P450-Dependent Monooxygenases. *Annu Rev Plant Physiol Plant Mol Biol* 49:311-343. doi:10.1146/annurev.arplant.49.1.311
37. Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. *BMC bioinformatics* 11:203. doi:10.1186/1471-2105-11-203
38. Chen T, Li E (2004) Structure and function of eukaryotic DNA methyltransferases. *Current topics in developmental biology* 60:55-89. doi:10.1016/S0070-2153(04)60003-2
39. Chen Z, Martinez DA, Gujja S, Sykes SM, Zeng Q, Szaniszlo PJ, Wang Z, Cuomo CA (2014) Comparative genomic and transcriptomic analysis of *wangiella dermatitidis*, a major cause of phaeohyphomycosis and a model black yeast human pathogen. *G3 (Bethesda)* 4 (4):561-578. doi:10.1534/g3.113.009241
40. Chevreux B (2005) MIRA: an Automated Genome and EST Assembler.
41. Chu HT, Hsiao WW, Chen JC, Yeh TJ, Tsai MH, Lin H, Liu YW, Lee SA, Chen CC, Tsao TT, Kao CY (2013) EBARDenovo: highly accurate de novo assembly of RNA-Seq with efficient chimera-detection. *Bioinformatics* 29 (8):1004-1010. doi:10.1093/bioinformatics/btt092
42. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009)

- Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4 (4):265-270. doi:10.1038/nnano.2009.12
43. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods* 5 (7):613-619. doi:10.1038/nmeth.1223
 44. Cogoni C, Macino G (1997) Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. *Proceedings of the National Academy of Sciences of the United States of America* 94 (19):10233-10238
 45. Cogoni C, Macino G (1999) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* 399 (6732):166-169. doi:10.1038/20215
 46. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J, Taga M, White GJ, Zhou S, Schwartz DC, Freitag M, Ma LJ, Danchin EG, Henrissat B, Coutinho PM, Nelson DR, Straney D, Napoli CA, Barker BM, Gribskov M, Rep M, Kroken S, Molnar I, Rensing C, Kennell JC, Zamora J, Farman ML, Selker EU, Salamov A, Shapiro H, Pangilinan J, Lindquist E, Lamers C, Grigoriev IV, Geiser DM, Covert SF, Temporini E, Vanetten HD (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS genetics* 5 (8):e1000618. doi:10.1371/journal.pgen.1000618
 47. Collins CR, Farrar JF (1978) STRUCTURAL RESISTANCES TO MASS TRANSFER IN THE LICHEN XANTHORIA PARIETINA. *New Phytologist* 81 (1):71-83. doi:10.1111/j.1469-8137.1978.tb01605.x
 48. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome biology* 17:13. doi:10.1186/s13059-016-0881-8
 49. Cooper DN (2002) Galectinomics: finding themes in complexity. *Biochimica et biophysica acta* 1572 (2-3):209-231
 50. Cummings RD, Liu FT (2009) Galectins. In: Varki A, Cummings RD, Esko JD et al. (eds) *Essentials of Glycobiology*. 2nd edn., Cold Spring Harbor (NY),

51. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes HW, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MI, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317 (5843):1400-1402. doi:10.1126/science.1143708
52. Dang Y, Yang Q, Xue Z, Liu Y (2011) RNA interference in fungi: pathways, functions, and applications. *Eukaryotic cell* 10 (9):1148-1155. doi:10.1128/EC.05109-11
53. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu J-R, Pan H, Read ND, Lee Y-H, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun M-H, Bohnert H, Coughlan S, Butler J, Calvo S, Ma L-J, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434:980. doi:10.1038/nature03449
54. Decker LM, Boone EC, Xiao H, Shanker BS, Boone SF, Kingston SL, Lee SA, Hammond TM, Shiu PK (2015) Complex formation of RNA silencing proteins in the perinuclear region of *Neurospora crassa*. *Genetics* 199 (4):1017-1021. doi:10.1534/genetics.115.174623
55. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW (2014) Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol* 10 (12):e1003998. doi:10.1371/journal.pcbi.1003998
56. Drew EA, Smith DC (1967) Studies in the Physiology of Lichens. VII. The Physiology of the Nostoc Symbiont of *Peltigera polydactyla* Compared with Cultured and Free-living Forms. *The New phytologist* 66 (3):379-388
57. Dyer PS (2002) Hydrophobins in the lichen symbiosis. *New Phytologist* 154 (1):1-4. doi:10.1046/j.1469-8137.2002.00387.x
58. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D,

- Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323 (5910):133-138. doi:10.1126/science.1162986
59. Elifio SL, De Lourdes C. C. Da Silva M, Iacomini M, Gorin PAJ (2000) A lectin from the lichenized Basidiomycete *Dictyonema glabratum*. *New Phytologist* 148 (2):327-334. doi:10.1046/j.1469-8137.2000.00758.x
 60. Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome research* 17 (1):69-73. doi:10.1101/gr.5145806
 61. Englund B (1977) The Physiology of the Lichen *Peltigera aphthosa*, with Special Reference to the Blue-green Phycobiont (*Nostoc* sp.). *Physiologia Plantarum* 41 (4):298-304. doi:10.1111/j.1399-3054.1977.tb04887.x
 62. Faugeron G, Rhounim L, Rossignol JL (1990) How does the cell count the number of ectopic copies of a gene in the premeiotic inactivation process acting in *Ascobolus immersus*? *Genetics* 124 (3):585-591
 63. Fincham JR, Connerton IF, Notarianni E, Harrington K (1989) Premeiotic disruption of duplicated and triplicated copies of the *Neurospora crassa* am (glutamate dehydrogenase) gene. *Current genetics* 15 (5):327-334
 64. Finotello F, Lavezzo E, Bianco L, Barzon L, Mazzon P, Fontana P, Toppo S, Di Camillo B (2014) Reducing bias in RNA sequencing data: a novel approach to compute counts. *BMC bioinformatics* 15 Suppl 1:S7. doi:10.1186/1471-2105-15-S1-S7
 65. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods* 7 (6):461-465. doi:10.1038/nmeth.1459
 66. Fontaniella B, Millanes AM, Vicente C, Legaz ME (2004) Concanavalin A binds to a mannose-containing ligand in the cell wall of some lichen phycobionts. *Plant physiology and*

- biochemistry : PPB / Societe francaise de physiologie vegetale 42 (10):773-779. doi:10.1016/j.plaphy.2004.09.003
67. Foss HM, Roberts CJ, Claeys KM, Selker EU (1993) Abnormal chromosome behavior in *Neurospora* mutants defective in DNA methylation. *Science* 262 (5140):1737-1741
 68. Francis WR, Christianson LM, Kiko R, Powers ML, Shaner NC, Haddock SH (2013) A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics* 14:167. doi:10.1186/1471-2164-14-167
 69. Freitag M, Hickey PC, Khlafallah TK, Read ND, Selker EU (2004) HP1 is essential for DNA methylation in *neurospora*. *Molecular cell* 13 (3):427-434
 70. Freitag M, Williams RL, Kothe GO, Selker EU (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proceedings of the National Academy of Sciences* 99 (13):8802-8807. doi:10.1073/pnas.132212899
 71. Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Dore LC, Weng X, Ji Q, Mets L, He C (2015) N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* 161 (4):879-892. doi:10.1016/j.cell.2015.04.010
 72. Fukushima A, Kusano M (2014) A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. *Journal of experimental botany* 65 (19):5619-5630. doi:10.1093/jxb/eru322
 73. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B

- (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422 (6934):859-868.
doi:10.1038/nature01554
74. Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. *Trends in genetics : TIG* 20 (9):417-423.
doi:10.1016/j.tig.2004.07.007
 75. Galun M (1998) Carbon Metabolism, In *Handbook of Lichenology*, , vol 1. Boca Raton: CRC Press,
 76. Galun M, Braun A, Frensdorff A, Galun E (1976) Hyphal walls of isolated lichen fungi: autoradiographic localization of precursor incorporation and binding of fluorescein-conjugated lectins. *Archives of microbiology* 108 (1):9-16
 77. Garg N, Zeng Y, Edlund A, Melnik AV, Sanchez LM, Mohimani H, Gurevich A, Miao V, Schiffler S, Lim YW, Luzzatto-Knaan T, Cai S, Rohwer F, Pevzner PA, Cichewicz RH, Alexandrov T, Dorrestein PC (2016) Spatial Molecular Architecture of the Microbial Community of a *Peltigera* Lichen. *mSystems* 1 (6).
doi:10.1128/mSystems.00139-16
 78. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj PK, Rani A, Singh RS, Raizada J, Singh K, Kumar S (2011) An RNA isolation system for plant tissues rich in secondary metabolites. *BMC research notes* 4:85. doi:10.1186/1756-0500-4-85
 79. Giannino D, Mele G, Cozza R, Bruno L, Testone G, Ticconi C, Frugis G, Bitonti MB, Innocenti AM, Mariotti D (2003) Isolation and characterization of a maintenance DNA-methyltransferase gene from peach (*Prunus persica* [L.] Batsch): transcript localization in vegetative and reproductive meristems of triple buds. *Journal of experimental botany* 54 (393):2623-2633.
doi:10.1093/jxb/erg292
 80. Gladyshev E (2017) Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. *Microbiol Spectr* 5 (4).
doi:10.1128/microbiolspec.FUNK-0042-2017
 81. Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128 (4):635-638.
doi:10.1016/j.cell.2007.02.006
 82. Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* 74:481-514.
doi:10.1146/annurev.biochem.74.010904.153721
 83. Gonzalez I, Ayuso-Sacido A, Anderson A, Genilloud O (2005) Actinomycetes isolated from lichens: evaluation of their diversity and detection of biosynthetic gene sequences. *FEMS Microbiol*

- Ecol 54 (3):401-415. doi:10.1016/j.femsec.2005.05.004
84. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research* 36 (10):3420-3435. doi:10.1093/nar/gkn176
 85. Gowher H, Ehrlich KC, Jeltsch A (2001) DNA from *Aspergillus flavus* contains 5-methylcytosine. *FEMS microbiology letters* 205 (1):151-155
 86. Goyon C, Faugeron G (1989) Targeted transformation of *Ascombolus immersus* and de novo methylation of the resulting duplicated DNA sequences. *Molecular and cellular biology* 9 (7):2818-2827
 87. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29 (7):644-U130. doi:Doi 10.1038/Nbt.1883
 88. Graham LE, Trest MT, Will-Wolf S, Miicke NS, Atonio LM, Piotrowski MJ, Knack JJ (2018) Microscopic and Metagenomic Analyses of *Peltigera ponojensis* (Peltigerales, Ascomycota). *Int J Plant Sci* 179 (3):241-255. doi:10.1086/696534
 89. Graia F, Lespinet O, Rimbault B, Dequard-Chablat M, Coppin E, Picard M (2001) Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. *Molecular microbiology* 40 (3):586-595
 90. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otilar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic acids research* 40 (Database issue):D26-32. doi:10.1093/nar/gkr947
 91. Grube M, Cardinale M, de Castro JV, Jr., Muller H, Berg G (2009) Species-specific structural and functional diversity of bacterial communities in lichen symbioses. *ISME J* 3 (9):1105-1115. doi:10.1038/ismej.2009.63
 92. Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK, Jr., Maiti R, Chan AP, Yu C, Farzad M, Wu D, White O, Town CD (2005) Complete reannotation of the *Arabidopsis* genome:

- methods, tools, protocols and the final release. *BMC Biol* 3:7. doi:10.1186/1741-7007-3-7
93. Hahn M, Viaud M, van Kan J (2014) The Genome of *Botrytis cinerea*, a Ubiquitous Broad Host Range Necrotroph. In: Dean RA, Lichens-Park A, Kole C (eds) *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 19-44. doi:10.1007/978-3-662-44056-8_2
 94. Hammond TM, Bok JW, Andrews MD, Reyes-Dominguez Y, Scazzocchio C, Keller NP (2008) RNA silencing gene truncation in the filamentous fungus *Aspergillus nidulans*. *Eukaryotic cell* 7 (2):339-349. doi:10.1128/EC.00355-07
 95. Hammond TM, Xiao H, Boone EC, Perdue TD, Pukkila PJ, Shiu PK (2011) SAD-3, a Putative Helicase Required for Meiotic Silencing by Unpaired DNA, Interacts with Other Components of the Silencing Machinery. *G3 (Bethesda)* 1 (5):369-376. doi:10.1534/g3.111.000570
 96. Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP (2015) Repeat-Induced Point Mutation: A Fungal-Specific, Endogenous Mutagenesis Process. In: van den Berg MA, Maruthachalam K (eds) *Genetic Transformation Systems in Fungi, Volume 2*. Springer International Publishing, Cham, pp 55-68. doi:10.1007/978-3-319-10503-1_4
 97. Harris EY, Ponts N, Levchuk A, Roch KL, Lonardi S (2010) BRAT: bisulfite-treated reads analysis tool. *Bioinformatics* 26 (4):572-573. doi:10.1093/bioinformatics/btp706
 98. Hill DJ, Smith DC (1972) Lichen physiology xii. The 'inhibition technique'. *New Phytologist* 71 (1):15-30. doi:10.1111/j.1469-8137.1972.tb04806.x
 99. Hodkinson BP, Lutzoni F (2009) A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. *Symbiosis* 49 (3):163-180. doi:10.1007/s13199-009-0049-3
 100. Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* 187 (4173):226-232
 101. Honda S, Bicocca VT, Gessaman JD, Rountree MR, Yokoyama A, Yu EY, Selker JM, Selker EU (2016) Dual chromatin recognition by the histone deacetylase complex HCHC is required for proper DNA methylation in *Neurospora crassa*. *Proceedings of the National Academy of Sciences of the United States of America* 113 (41):E6135-E6144. doi:10.1073/pnas.1614279113
 102. Honda S, Lewis ZA, Huarte M, Cho LY, David LL, Shi Y, Selker

- EU (2010) The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. *Genes & development* 24 (5):443-454. doi:10.1101/gad.1893210
103. Honda S, Lewis ZA, Shimada K, Fischle W, Sack R, Selker EU (2012) Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. *Nat Struct Mol Biol* 19 (5):471-477, S471. doi:10.1038/nsmb.2274
 104. Honda S, Selker EU (2008) Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in *Neurospora crassa*. *Molecular and cellular biology* 28 (19):6044-6055. doi:10.1128/MCB.00823-08
 105. Honegger R (1984) Ultrastructural studies on conidiomata, conidiophores, and conidiogenous cells in six lichen-forming Ascomycetes. *Canadian Journal of Botany* 62 (10):2081-2093. doi:10.1139/b84-283
 106. Honegger R (1996) *Mycobionts, vol III. Lichen Biology*. Cambridge University Press, Cambridge, pp. 27-39,
 107. Howe ML, Barrett JT (1970) Studies on a hemagglutinin from the lichen *Parmelia michauxiana*. *Biochimica et biophysica acta* 215 (1):97-104
 108. Idnurm A, Howlett BJ (2003) Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. *Fungal genetics and biology : FG & B* 39 (1):31-37
 109. Ikeda K, Nakayashiki H, Kataoka T, Tamba H, Hashimoto Y, Tosa Y, Mayama S (2002) Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Molecular microbiology* 45 (5):1355-1364
 110. Irelan JT, Selker EU (1996) Gene silencing in filamentous fungi: RIP, MIP and quelling. *Journal of Genetics* 75 (3):313-324. doi:10.1007/bf02966311
 111. Jacobson ES (2000) Pathogenic roles for fungal melanins. *Clin Microbiol Rev* 13 (4):708-717
 112. Jensen M, Feige GB, Waterkotte A (1991) Mannitol-1-Phosphate Dehydrogenase in *Pseudevernia Furfuracea*. *The Lichenologist* 23 (02):187-196
 113. Jeon J, Choi J, Lee GW, Park SY, Huh A, Dean RA, Lee YH (2015) Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Scientific reports*

- 5:8567. doi:10.1038/srep08567
114. Jones PA, Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20 (1):85-93
 115. Joneson S, Armaleo D, Lutzoni F (2011) Fungal and algal gene expression in early developmental stages of lichen-symbiosis. *Mycologia* 103 (2):291-306. doi:10.3852/10-064
 116. Junttila S, Rudd S (2012) Characterization of a transcriptome from a non-model organism, *Cladonia rangiferina*, the grey reindeer lichen, using high-throughput next generation sequencing and EST sequence data. *BMC Genomics* 13 (1):575
 117. Kampa A, Gagunashvili AN, Gulder TA, Morinaka BI, Daolio C, Godejohann M, Miao VP, Piel J, Andresson O (2013) Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proceedings of the National Academy of Sciences of the United States of America* 110 (33):E3129-3137. doi:10.1073/pnas.1305867110
 118. Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of molecular biology* 428 (4):726-731. doi:10.1016/j.jmb.2015.11.006
 119. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research* 22 (12):2497-2506. doi:10.1101/gr.143008.112
 120. Kershaw MJ, Talbot NJ (1998) Hydrophobins and repellents: proteins with fundamental roles in fungal morphogenesis. *Fungal genetics and biology : FG & B* 23 (1):18-33. doi:10.1006/fgbi.1997.1022
 121. Khosravi C, Benocci T, Battaglia E, Benoit I, de Vries RP (2015) Chapter One - Sugar Catabolism in *Aspergillus* and Other Fungi Related to the Utilization of Plant Biomass. In: Sariaslani S, Gadd GM (eds) *Advances in Applied Microbiology*, vol 90. Academic Press, pp 1-28. doi:<https://doi.org/10.1016/bs.aambs.2014.09.005>
 122. Klocko AD, Rountree MR, Grisafi PL, Hays SM, Adhvaryu KK, Selker EU (2015) *Neurospora* importin alpha is required for normal heterochromatic formation and DNA methylation. *PLoS genetics* 11 (3):e1005083. doi:10.1371/journal.pgen.1005083
 123. Kouzminova E, Selker EU (2001) *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *The EMBO journal* 20 (15):4309-4323.

- doi:10.1093/emboj/20.15.4309
124. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27 (11):1571-1572. doi:10.1093/bioinformatics/btr167
 125. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG (1994) The DNA (cytosine-5) methyltransferases. *Nucleic acids research* 22 (1):1-10
 126. Lange OL, Green TG, Reichenberger H, Hesbacher S, Proksch P (1997) Do secondary substances in the thallus of a lichen promote CO₂ diffusion and prevent depression of net photosynthesis at high water content? *Oecologia* 112 (1):1-3. doi:10.1007/s004420050275
 127. Langmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome biology* 11 (8):R83. doi:10.1186/gb-2010-11-8-r83
 128. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9 (4):357-359. doi:10.1038/nmeth.1923
 129. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10 (3):R25. doi:10.1186/gb-2009-10-3-r25
 130. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11 (3):204-220. doi:10.1038/nrg2719
 131. Lawrey JD, Diederich P (2003) Lichenicolous Fungi: Interactions, Evolution, and Biodiversity. *The Bryologist* 106 (1):80-120
 132. Lee DW, Freitag M, Selker EU, Aramayo R (2008) A cytosine methyltransferase homologue is essential for sexual development in *Aspergillus nidulans*. *PloS one* 3 (6):e2531. doi:10.1371/journal.pone.0002531
 133. Lee DW, Pratt RJ, McLaughlin M, Aramayo R (2003) An argonaute-like protein is required for meiotic silencing. *Genetics* 164 (2):821-828
 134. Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz M (2016) Third-generation sequencing and the future of genomics. *bioRxiv*
 135. Leffler H, Carlsson S, Hedlund M, Qian Y, Poirier F (2004) Introduction to galectins. *Glycoconjugate journal* 19 (7-9):433-440. doi:10.1023/B:GLYC.0000014072.34840.04
 136. Lehr H, Fleminger G, Galun M (1995) Lectin from the lichen

- Peltigera membranacea* Ach. Nyl.: characterization and function. *Symbiosis Rehovot* 18 (1):1-13
137. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, Selker EU (2010) DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC. *PLoS genetics* 6 (11):e1001196. doi:10.1371/journal.pgen.1001196
 138. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12:323. doi:10.1186/1471-2105-12-323
 139. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26 (5):589-595. doi:10.1093/bioinformatics/btp698
 140. Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J, Terzaghi W, Li S, Deng XW (2008) High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell* 20 (2):259-276. doi:10.1105/tpc.107.056879
 141. Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K (2011) RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics* 12:540. doi:10.1186/1471-2164-12-540
 142. Liba CM, Ferrara FI, Manfio GP, Fantinatti-Garboggini F, Albuquerque RC, Pavan C, Ramos PL, Moreira-Filho CA, Barbosa HR (2006) Nitrogen-fixing chemo-organotrophic bacteria isolated from cyanobacteria-deprived lichens and their ability to solubilize phosphate and to release amino acids and phytohormones. *J Appl Microbiol* 101 (5):1076-1086. doi:10.1111/j.1365-2672.2006.03010.x
 143. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133 (3):523-536. doi:10.1016/j.cell.2008.03.029
 144. Liu SY, Lin JQ, Wu HL, Wang CC, Huang SJ, Luo YF, Sun JH, Zhou JX, Yan SJ, He JG, Wang J, He ZM (2012) Bisulfite sequencing reveals that *Aspergillus flavus* holds a hollow in DNA methylation. *PloS one* 7 (1):e30349. doi:10.1371/journal.pone.0030349
 145. Lockhart CM, Rowell P, Stewart WDP (1978) Phytohaemagglutinins from the nitrogen-fixing lichens *Peltigera canina* and *P. polydactyla*. *FEMS microbiology letters* 3 (3):127-

130. doi:<http://dx.doi.org/>
146. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15 (12):550. doi:10.1186/s13059-014-0550-8
147. Lugones LG, Bosscher JS, Scholtmeyer K, de Vries OM, Wessels JG (1996) An abundant hydrophobin (ABH1) forms hydrophobic rodlet layers in *Agaricus bisporus* fruiting bodies. *Microbiology* 142 (Pt 5):1321-1329. doi:10.1099/13500872-142-5-1321
148. Luyten E, Swinnen E, Vlassak K, Verreth C, Dombrecht B, Vanderleyden J (2001) Analysis of a symbiosis-specific cytochrome P450 homolog in *Rhizobium* sp. BR816. *Mol Plant Microbe Interact* 14 (7):918-924. doi:10.1094/MPMI.2001.14.7.918
149. Lyimo B, Yagi F, Minami Y (2011) Primary structure and specificity of a new member of galectin family from the Amethyst deceiver mushroom *Laccaria amethystina*. *Bioscience, biotechnology, and biochemistry* 75 (1):62-69. doi:10.1271/bbb.100542
150. MacFarlane JD, Kershaw KA (1985) Some Aspects of Carbohydrate Metabolism in Lichens. In: Brown DH (ed) *Lichen Physiology and Cell Biology*. Springer US, Boston, MA, pp 1-8. doi:10.1007/978-1-4613-2527-7_1
151. Malagnac F, Silar P (2011) Chapter 13 - Epigenetics of Eukaryotic Microbes A2 - Tollefsbol, Trygve. In: *Handbook of Epigenetics*. Academic Press, San Diego, pp 185-201. doi:<https://doi.org/10.1016/B978-0-12-375709-8.00013-7>
152. Malagnac F, Wendel B, Goyon C, Faugeron G, Zickler D, Rossignol JL, Noyer-Weidner M, Vollmayr P, Trautner TA, Walter J (1997) A gene essential for de novo methylation and development in *Ascobolus* reveals a novel type of eukaryotic DNA methyltransferase structure. *Cell* 91 (2):281-290
153. Mangul S, Caciula A, Al Seesi S, Brinza D, Mndoiu I, Zelikovsky A (2014) Transcriptome assembly and quantification from Ion Torrent RNA-Seq data. *BMC Genomics* 15 Suppl 5:S7. doi:10.1186/1471-2164-15-S5-S7
154. Manoharan SS, Miao VP, Andresson OS (2012) LEC-2, a highly variable lectin in the lichen. *Symbiosis* 58 (1-3):91-98. doi:10.1007/s13199-012-0206-y
155. Margolin BS, Garrett-Engle PW, Stevens JN, Fritz DY, Garrett-Engle C, Metzenberg RL, Selker EU (1998) A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element

- inactivated by repeat-induced point mutation. *Genetics* 149 (4):1787-1797
156. Marini AM, Soussi-Boudekou S, Vissers S, Andre B (1997) A family of ammonium transporters in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 17 (8):4282-4293
 157. Mathis R, Van Gijsegem F, De Rycke R, D'Haese W, Van Maelsaeke E, Anthonio E, Van Montagu M, Holsters M, Vereecke D (2005) Lipopolysaccharides as a communication signal for progression of legume endosymbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 102 (7):2655-2660. doi:10.1073/pnas.0409816102
 158. Mato JM, Alvarez L, Ortiz P, Pajares MA (1997) S-adenosylmethionine synthesis: molecular mechanisms and clinical implications. *Pharmacology & therapeutics* 73 (3):265-280
 159. Mautino MR, Rosa AL (1998) Analysis of models involving enzymatic activities for the occurrence of C-->T transition mutations during repeat-induced point mutation (RIP) in *Neurospora crassa*. *Journal of theoretical biology* 192 (1):61-71. doi:10.1006/jtbi.1997.0608
 160. McDonald T (2011) Genomic Insights Into the Lichen Symbiosis: *Cladonia grayi*. Dissertation, Duke University,
 161. McDonald TR, Dietrich FS, Lutzoni F (2012) Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes: toward a new functional and evolutionary classification. *Molecular biology and evolution* 29 (1):51-60. doi:10.1093/molbev/msr123
 162. Meena M, Prasad V, Zehra A, Gupta VK, Upadhyay RS (2015) Mannitol metabolism during pathogenic fungal-host interactions under stressed conditions. *Front Microbiol* 6:1019. doi:10.3389/fmicb.2015.01019
 163. Miadlikowska J, Lutzoni F (2004) Phylogenetic classification of peltigeralean fungi (Peltigerales, Ascomycota) based on ribosomal RNA small and large subunits. *American journal of botany* 91 (3):449-464. doi:Doi 10.3732/Ajb.91.3.449
 164. Miao VPW, Manoharan SS, Snaebjarnarson V, Andresson OS (2012) Expression of lec-1, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea*. *Symbiosis* 57 (1):23-31. doi:Doi 10.1007/S13199-012-0175-1
 165. Millanes AM, Diederich P, Wedin M (2016) *Cyphobasidium* gen. nov., a new lichen-inhabiting lineage in the Cystobasidiomycetes (Pucciniomycotina, Basidiomycota, Fungi). *Fungal biology* 120

- (11):1468-1477. doi:10.1016/j.funbio.2015.12.003
166. Millbank JW, Kershaw KA (1969) NITROGEN METABOLISM IN LICHENS. *New Phytologist* 68 (3):721-729. doi:10.1111/j.1469-8137.1969.tb06476.x
 167. Mishra PK, Baum M, Carbon J (2011) DNA methylation regulates phenotype-dependent transcriptional activity in *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America* 108 (29):11965-11970. doi:10.1073/pnas.1109631108
 168. Molina MC, Stocker-Worgotter E, Turk R, Bajon C, Vicente C (1998) Secreted, glycosylated arginase from *Xanthoria parietina* thallus induces loss of cytoplasmic material from *Xanthoria* photobionts. *Cell adhesion and communication* 6 (6):481-490
 169. Molina MC, Vincente C (1995) Correlations between enzymatic activity of lectins, putrescine content and chloroplast damage in *Xanthoria parietina* phycobionts. *Cell adhesion and communication* 3 (1):1-12
 170. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A, Salamov A, Ahrendt SR, Lau R, Bowen BP, Lipzen A, Sullivan W, Andreopoulos BB, Clum A, Lindquist E, Daum C, Northen TR, Kunde-Ramamoorthy G, Schmitz RJ, Gryganskyi A, Culley D, Magnuson J, James TY, O'Malley MA, Stajich JE, Spatafora JW, Visel A, Grigoriev IV (2017) Widespread adenine N6-methylation of active genes in fungi. *Nature genetics* 49 (6):964-968. doi:10.1038/ng.3859
 171. Montanini B, Chen PY, Morselli M, Jaroszewicz A, Lopez D, Martin F, Ottonello S, Pellegrini M (2014) Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. *Genome biology* 15 (7):411. doi:10.1186/s13059-014-0411-5
 172. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5 (7):621-628. doi:10.1038/nmeth.1226
 173. Munroe DJ, Harris TJ (2010) Third-generation sequencing fireworks at Marco Island. *Nature biotechnology* 28 (5):426-428. doi:10.1038/nbt0510-426
 174. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320 (5881):1344-

1349. doi:10.1126/science.1158441
175. Nakayashiki H, Kadotani N, Mayama S (2006) Evolution and diversification of RNA silencing proteins in fungi. *Journal of molecular evolution* 63 (1):127-135. doi:10.1007/s00239-005-0257-2
 176. Nash I, T. (Ed.). (2008) *Lichen Biology*. 2 edn. Cambridge University Press, Cambridge. doi:DOI: 10.1017/CBO9780511790478
 177. Neu E, Featherston J, Rees J, Debener T (2017) A draft genome sequence of the rose black spot fungus *Diplocarpon rosae* reveals a high degree of genome duplication. *PloS one* 12 (10):e0185310. doi:10.1371/journal.pone.0185310
 178. Ninnemann O, Jauniaux JC, Frommer WB (1994) Identification of a high affinity NH₄⁺ transporter from plants. *The EMBO journal* 13 (15):3464-3471
 179. Nosanchuk JD, Casadevall A (2003) The contribution of melanin to microbial pathogenesis. *Cell Microbiol* 5 (4):203-223
 180. Nowrousian M, Cebula P (2005) The gene for a lectin-like protein is transcriptionally activated during sexual development, but is not essential for fruiting body formation in the filamentous fungus *Sordaria macrospora*. *BMC Microbiology* 5:64
 181. Nowrousian M, Stajich JE, Chu M, Engh I, Espagne E, Halliday K, Kamerewerd J, Kempken F, Knab B, Kuo HC, Osiewacz HD, Poggeler S, Read ND, Seiler S, Smith KM, Zickler D, Kuck U, Freitag M (2010) De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. *PLoS genetics* 6 (4):e1000891. doi:10.1371/journal.pgen.1000891
 182. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99 (3):247-257
 183. Oliver RP, Hane JK (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC bioinformatics* 9. doi:10.1186/1471-2105-9-478
 184. Oszolak F, Milos PM (2011) Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdiscip Rev RNA* 2 (4):565-570. doi:10.1002/wrna.84
 185. Pal K, van Diepeningen AD, Varga J, Hoekstra RF, Dyer PS, Debets AJ (2007) Sexual and vegetative compatibility genes in the aspergilli. *Studies in mycology* 59:19-30. doi:10.3114/sim.2007.59.03

186. Paoletti M, Clavé C (2007) The Fungus-Specific HET Domain Mediates Programmed Cell Death in *Podospira anserina*. *Eukaryotic cell* 6 (11):2001-2008. doi:10.1128/ec.00129-07
187. Park SY, Choi J, Lee GW, Kim JA, Oh SO, Jeong MH, Yu NH, Kim S, Lee YH, Hur JS (2014) Draft Genome Sequence of Lichen-Forming Fungus *Cladonia metacorallifera* Strain KoLRI002260. *Genome announcements* 2 (1). doi:10.1128/genomeA.01065-13
188. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23 (9):1061-1067. doi:10.1093/bioinformatics/btm071
189. Patel TK, Williamson JD (2016) Mannitol in Plants, Fungi, and Plant–Fungal Interactions. *Trends in Plant Science* 21 (6):486-497. doi:<https://doi.org/10.1016/j.tplants.2016.01.006>
190. Piel J (2010) 2.14 - The Chemistry of Symbiotic Interactions. In: *Comprehensive Natural Products II*. Elsevier, Oxford, pp 475-510. doi:<https://doi.org/10.1016/B978-008045382-8.00049-6>
191. Piper PW, Ortiz-Calderon C, Holyoak C, Coote P, Cole M (1997) Hsp30, the integral plasma membrane heat shock protein of *Saccharomyces cerevisiae*, is a stress-inducible regulator of plasma membrane H(+)-ATPase. *Cell Stress Chaperones* 2 (1):12-24
192. Proffitt JH, Davie JR, Swinton D, Hattman S (1984) 5-Methyleytosine is not detectable in *Saccharomyces cerevisiae* DNA. *Molecular and cellular biology* 4 (5):985-988
193. Rai AN, Rowell P, Stewart WDP (1980) NH₄⁺ ASSIMILATION AND NITROGENASE REGULATION IN THE LICHEN *PELTIGERA APHTHOSA* WILLD. *New Phytologist* 85 (4):545-555. doi:10.1111/j.1469-8137.1980.tb00770.x
194. Raine A, Liljedahl U, Nordlund J (2018) Data quality of whole genome bisulfite sequencing on Illumina platforms. *PloS one* 13 (4):e0195972. doi:10.1371/journal.pone.0195972
195. Ramirez ML, Chulze SN, Magan N (2004) Impact of osmotic and matric water stress on germination, growth, mycelial water potentials and endogenous accumulation of sugars and sugar alcohols in *Fusarium graminearum*. *Mycologia* 96 (3):470-478
196. Rhounim L, Rossignol JL, Faugeron G (1992) Epimutation of repeated genes in *Ascobolus immersus*. *The EMBO journal* 11 (12):4451-4457
197. Richardson DHS (1992) Pollution Monitoring with Lichens.

- Richmond Publishing, Slough, England
198. Richardson DHS, Hill DJ, Smith DC (1968) Lichen Physiology. XI. The Role of the Alga in Determining the Pattern of Carbohydrate Movement Between Lichen Symbionts. *The New phytologist* 67 (3):469-486
 199. Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics* 14 (1):9-25
 200. Roberts RJ (1978) Restriction and modification enzymes and their recognition sequences. *Gene* 4 (3):183-194
 201. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1):139-140. doi:10.1093/bioinformatics/btp616
 202. Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Molecular microbiology* 6 (22):3343-3353
 203. Rosa AL, Folco HD, Mautino MR (2004) In vivo levels of S-adenosylmethionine modulate C:G to T:A mutations associated with repeat-induced point mutation in *Neurospora crassa*. *Mutation research* 548 (1-2):85-95. doi:10.1016/j.mrfmmm.2004.01.001
 204. Rosic NN, Pernice M, Dunn S, Dove S, Hoegh-Guldberg O (2010) Differential regulation by heat stress of novel cytochrome P450 genes from the dinoflagellate symbionts of reef-building corals. *Appl Environ Microbiol* 76 (9):2823-2829. doi:10.1128/AEM.02984-09
 205. Rossignol J-L, Picard M (1991) *Ascobolus immersus* and *Podospira anserina*: Sex, Recombination, Silencing, and Death A2 - Bennett, J.W. In: Lasure LL (ed) *More Gene Manipulations in Fungi*. Academic Press, San Diego, pp 266-290. doi:<http://dx.doi.org/10.1016/B978-0-12-088642-5.50019-3>
 206. Rowell P, Rai AN, Stewart WDP (1985) Studies on the Nitrogen Metabolism of the Lichens *Peltigera Aphthosa* and *Peltigera Canina*. In: Brown DH (ed) *Lichen Physiology and Cell Biology*. Springer US, Boston, MA, pp 145-160. doi:10.1007/978-1-4613-2527-7_10
 207. Sacristan M, Millanes AM, Legaz ME, Vicente C (2006) A Lichen Lectin Specifically Binds to the alpha-1,4-Polygalactoside Moiety of Urease Located in the Cell Wall of Homologous Algae. *Plant signaling & behavior* 1 (1):23-27
 208. Salem M, Paneru B, Al-Tobasei R, Abdouni F, Thorgaard GH,

- Rexroad CE, Yao J (2015) Transcriptome assembly, gene annotation and tissue gene expression atlas of the rainbow trout. *PloS one* 10 (3):e0121778. doi:10.1371/journal.pone.0121778
209. Sanders WB (2001) Lichens: The Interface between Mycology and Plant Morphology. *BioScience* 51 (12):1025-1036. doi:10.1641/0006-3568(2001)051[1025:LTIBMA]2.0.CO;2
 210. Schmitt I, Lumbsch HT (2009) Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. *PloS one* 4 (2):e4437. doi:10.1371/journal.pone.0004437
 211. Seidl MF (2017) Adenine N6-methylation in diverse fungi. *Nature genetics* 49 (6):823-824. doi:10.1038/ng.3873
 212. Selker EU (1990) Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annual review of genetics* 24:579-613. doi:10.1146/annurev.ge.24.120190.003051
 213. Sendler E, Johnson GD, Krawetz SA (2011) Local and global factors affecting RNA sequencing analysis. *Analytical biochemistry* 419 (2):317-322. doi:10.1016/j.ab.2011.08.013
 214. Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. *Trends in ecology & evolution* 23 (10):578-587. doi:10.1016/j.tree.2008.06.005
 215. Sharon N (2007) Lectins: carbohydrate-specific reagents and biological recognition molecules. *The Journal of biological chemistry* 282 (5):2753-2764. doi:10.1074/jbc.X600004200
 216. Shiu PK, Raju NB, Zickler D, Metzenberg RL (2001) Meiotic silencing by unpaired DNA. *Cell* 107 (7):905-916
 217. Sigurbjornsdottir MA, Andresson OS, Vilhelmsson O (2015) Analysis of the *Peltigera membranacea* metagenome indicates that lichen-associated bacteria are involved in phosphate solubilization. *Microbiology* 161 (Pt 5):989-996. doi:10.1099/mic.0.000069
 218. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19):3210-3212. doi:10.1093/bioinformatics/btv351
 219. Singer MJ, Marcotte BA, Selker EU (1995) DNA Methylation Associated with Repeat-Induced Point Mutation in *Neurospora-Crassa*. *Molecular and cellular biology* 15 (10):5586-5597
 220. Singh RS, Bhari R, Kaur HP (2010) Mushroom lectins: current status and future perspectives. *Critical reviews in biotechnology* 30 (2):99-126. doi:10.3109/07388550903365048
 221. Singh RS, Bhari R, Kaur HP (2011) Current trends of lectins from

- microfungi. Critical reviews in biotechnology 31 (3):193-210.
doi:10.3109/07388551.2010.505911
222. Spribille T, Tuovinen V, Resl P, Vanderpool D, Wolinski H, Aime MC, Schneider K, Stabentheiner E, Toome-Heller M, Thor G, Mayrhofer H, Johannesson H, McCutcheon JP (2016) Basidiomycete yeasts in the cortex of ascomycete macrolichens. Science 353 (6298):488
 223. Steinhauser SS, Andresson OS, Palsson A, Werth S (2016) Fungal and cyanobacterial gene expression in a lichen symbiosis: Effect of temperature and location. Fungal biology 120 (10):1194-1208. doi:10.1016/j.funbio.2016.07.002
 224. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9 (6):465-476. doi:10.1038/nrg2341
 225. Talbot NJ, Kershaw MJ, Wakley GE, De Vries O, Wessels J, Hamer JE (1996) MPG1 Encodes a Fungal Hydrophobin Involved in Surface Interactions during Infection-Related Development of *Magnaporthe grisea*. Plant Cell 8 (6):985-999. doi:10.1105/tpc.8.6.985
 226. Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature 414 (6861):277-283. doi:10.1038/35104508
 227. Tamaru H, Zhang X, McMillen D, Singh PB, Nakayama J, Grewal SI, Allis CD, Cheng X, Selker EU (2003) Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. Nature genetics 34 (1):75-79. doi:10.1038/ng1143
 228. Tan L, Shi YG (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development 139 (11):1895-1902. doi:10.1242/dev.070771
 229. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22 (22):4673-4680
 230. Tiwari S, Thakur R, Shankar J (2015) Role of Heat-Shock Proteins in Cellular Function and in the Biology of Fungi. Biotechnology Research International 2015:11. doi:10.1155/2015/132635
 231. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (9):1105-1111. doi:10.1093/bioinformatics/btp120

232. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28 (5):511-515. doi:10.1038/nbt.1621
233. UniProt C (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic acids research* 38 (Database issue):D142-148. doi:10.1093/nar/gkp846
234. Vasta GR (2009) Roles of galectins in infection. *Nature reviews Microbiology* 7 (6):424-438. doi:10.1038/nrmicro2146
235. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular ecology* 17 (7):1636-1647. doi:10.1111/j.1365-294X.2008.03666.x
236. Wan ZY, Xia JH, Lin G, Wang L, Lin VC, Yue GH (2016) Genome-wide methylation analysis identified sexually dimorphic methylated regions in hybrid tilapia. *Scientific reports* 6:35903. doi:10.1038/srep35903
237. Wang Y, Zhang X, Zhou Q, Zhang X, Wei J (2015) Comparative transcriptome analysis of the lichen-forming fungus *Endocarpon pusillum* elucidates its drought adaptation mechanisms. *Sci China Life Sci* 58 (1):89-100. doi:10.1007/s11427-014-4760-9
238. Wang Y-Y, Liu B, Zhang X-Y, Zhou Q-M, Zhang T, Li H, Yu Y-F, Zhang X-L, Hao X-Y, Wang M, Wang L, Wei J-C (2014) Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genomics* 15 (1):34
239. Wassenegger M, Pelissier T (1998) A model for RNA-mediated gene silencing in higher plants. *Plant Mol Biol* 37 (2):349-362
240. Watters MK, Randall TA, Margolin BS, Selker EU, Stadler DR (1999) Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. *Genetics* 153 (2):705-714
241. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453 (7199):1239-1243. doi:10.1038/nature07002
242. Wilken PM, Steenkamp ET, van der Nest MA, Wingfield MJ, de Beer ZW, Wingfield BD (2018) Unexpected placement of the

- MAT1-1-2 gene in the MAT1-2 idiomorph of *Thielaviopsis*.
Fungal genetics and biology : FG & B 113:32-41.
doi:10.1016/j.fgb.2018.01.007
243. Wilkinson CR, Bartlett R, Nurse P, Bird AP (1995) The fission yeast gene *pmt1+* encodes a DNA methyltransferase homologue. *Nucleic acids research* 23 (2):203-210
 244. Wisecaver JH, Slot JC, Rokas A (2014) The evolution of fungal metabolic pathways. *PLoS genetics* 10 (12):e1004816.
doi:10.1371/journal.pgen.1004816
 245. Wosten HA, Schuren FH, Wessels JG (1994) Interfacial self-assembly of a hydrophobin into an amphipathic protein membrane mediates fungal attachment to hydrophobic surfaces. *The EMBO journal* 13 (24):5848-5854
 246. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M, Tackett A, Wang G, Hon LS, Fang G, Swenberg JA, Xiao AZ (2016) DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* 532 (7599):329-333. doi:10.1038/nature17640
 247. Xavier BB, Miao VP, Jonsson ZO, Andresson OS (2012) Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: features and phylogeny. *Fungal biology* 116 (7):802-814.
doi:10.1016/j.funbio.2012.04.013
 248. Xi Y, Li W (2009) BSMAP: whole genome bisulfite sequence MAPPING program. *BMC bioinformatics* 10 (1):232
 249. Yoshino K, Sonoda M, Sakamoto K, Yamamoto Y (2016) Candidates of ribitol transporter gene obtained from *Ramalina conduplicans*. Paper presented at the The 8th IAL Symposium, Helsinki, Finland, August 1-5,
 250. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328 (5980):916-919. doi:10.1126/science.1186366
 251. Zemach A, Zilberman D (2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current biology : CB* 20 (17):R780-785. doi:10.1016/j.cub.2010.07.007
 252. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J, Li C, Liu B, Luo Y, Zhu Y, Zhang N, He S, He C, Wang H, Chen D (2015) N6-methyladenine DNA modification in *Drosophila*. *Cell* 161 (4):893-906.
doi:10.1016/j.cell.2015.04.018
 253. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H,

- Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126 (6):1189-1201. doi:10.1016/j.cell.2006.08.003
254. Zhao QY, Wang Y, Kong YM, Luo D, Li X, Hao P (2011) Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC bioinformatics* 12 Suppl 14:S2. doi:10.1186/1471-2105-12-S14-S2
255. Zilberman D (2017) An evolutionary case for functional gene body methylation in plants and animals. *Genome biology* 18 (1):87. doi:10.1186/s13059-017-1230-2
256. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature genetics* 39 (1):61-69. doi:10.1038/ng1929.

APPENDIXES

Appendix A: *Peltigera membranacea* sample collection, DNA isolation and genome features.

1. *Peltigera membranacea* sample collection.

P. membranacea (accession XBB013, LA-31632 (IINH)) was sampled at Keldnagil, Reykjavik, Iceland within a 12-m span (coordinates 41 64.70 N, 21 46.60 W) (Xavier et al. 2012). Both apothecia and thallus tissues were collected separately into sterile containers. Prior to genomic DNA isolation, samples were extensively cleaned under running tap water followed by distilled water and examined under a stereomicroscope to remove any surface debris. The DNA samples were then sequenced via the Roche 454 technology. Approximately, 7.39 and 7.88 gigabases of 454 raw data were obtained for apothecia and thallus tissues, respectively.

2. *Peltigera membranacea* DNA isolation protocol.

1. Collect and clean the samples.
2. Set the water bath at 65°C.
3. Grind the sample (2gm) with pestle to a powder with liquid Nitrogen in a pre-chilled mortar. (mortar + liquid nitrogen + add samples, wait till the Nitrogen stops churning + then start homogenizing)
4. Using a pre-chilled metal spatula to scrape the powdered tissue to the premeasured 25ml Lysis buffer (Temperature of Buffer 65°C).
5. Incubate at 65°C for 20 mins. (Mix the contents of the tube during the incubation).
6. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
7. Transfer the supernatant to a new tube and add 8.56ml of 10.5M Ammonium acetate (i.e. 8.56 ml for 20ml & 535µl for 1.25ml)
8. Mix the contents and incubate on ice for 20 mins.
9. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
10. Transfer the supernatant to a new tube, discard the pellet and add 12.8ml of isopropanol (i.e. 12.8ml for 20ml & 800µl for 1.25ml)
11. Mix the contents and incubate on ice for 20 mins.
12. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
13. Discard the supernatant.
14. Add TE buffer and RNase A to the pellet. (1ml TE + 4µl of RNase A)
15. Incubate the tubes at 50°C for 20–60 mins till the pellet is dissolved. (gently tap the tube to dissolve the pellet during incubation).

16. Centrifuge briefly (short spin) and discard the supernatant.
17. Add equal volume of Phenol: Chloroform: Isoamyl alcohol (25:24:1).
18. Vortex briefly.
19. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
20. Transfer the aqueous phase to a fresh tube without disturbing the interphase.
21. Add equal volume of saturated chloroform.
22. Vortex briefly.
23. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
24. Transfer the aqueous phase to a fresh tube without disturbing the interphase.
25. Add 1/10th volume of 3M Sodium acetate.
26. Add equal volume of 100% ethanol.
27. Mix the contents and incubate on ice for 20 mins.
28. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
29. Wash the pellet with 70% ethanol.
30. Centrifuge at 5,000 rpm for 15mins or 14,000rpm for 4mins.
31. Air dry the pellet and add 50 µl of TE buffer to dissolve the pellet.
32. Store the isolated DNA at –20° C

3. *Peltigera membranacea* assembly.

The *P. membranacea* 454 DNA contigs (Pmem 454Allcontigs) consist of sequences from *Nostoc*, the mycobiont and assorted bacteria. Before scaffolding, the 454 DNA contigs were binned and sorted by mapping them to rhizine RNA-Seq reads to sort out the mycobiont contigs. Using MIRA v3 (Chevreux 2005), the mycobiont contigs (14,674) were assembled into 3,033 scaffolds (N50-34.52 kbp) with an estimated genome size of 43 Mbp.

Table S1: *Peltigera membranacea* genome statistics.

	454Allcontigs	Sorted mycobiont contigs	Pmem v1.1
Statistics for contig lengths:			
Min contig length:	100	100	1,014
Max contig length:	58,399	55267	244,130
Mean contig length:	899.92	2977.81	14227.74
Standard deviation of contig length:	1733.45	3853.31	21371.24
Median contig length:	486	1544	5,445
N50 contig length:	1,660	5970	34,520
Statistics for numbers of contigs:			
Number of contigs:	150,581	14674	3,033
Number of contigs >=1kb:	28,879	9087	3,033
Number of contigs in N50:	16,125	2081	351
Statistics for bases in the contigs:			
Number of bases in all contigs:	135,511,117	43696402	43,152,721
Number of bases in contigs >=1kb:	83,927,598	40847977	43,152,721
GC Content of contigs:	44.89%	0.4494	44.24%

Appendix B: Metagenomic analysis.

1. Bioinformatic processing

The raw 454 reads from apothecia and thallus whole samples were imported into CLC Genomics Workbench v9.5.3, and the sequences were trimmed using default parameters (trim using quality scores = 0.05 and trim ambiguous nucleotides=2) Table S2. The trimmed 454 reads were used as input, and the OTUs with 97% similarity were sorted using microbial genomics module implemented in CLC Genomics Workbench v9.5.3. Reference databases were downloaded from the UNITE (<https://unite.ut.ee/>) (v7.2) (Kõljalg et al. 2014) for ITS (fungi), and the SILVA database (v128) (Quast et al. 2013) for 16S, 18S and 28S rRNA. UNITE database contains 7,802 ITS sequences from 2,120 species and all fungal ITS sequences from the International Nucleotide Sequence Databases (INSD: NCBI, EMBL, DDBJ).

The 18S and 28S rRNA reads classified as OTUs using 97% similarity were assigned to taxonomical classification using NGS analysis pipeline of SILVA rRNA gene database project (SILVAngs 1.3) (Quast et al., 2013). Each read was aligned using 98% similarity using the SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN (revision 21008)) (Pruesse et al., 2012) against the SILVA SSU rRNA SEED and quality controlled (Quast et al., 2013). Reads shorter than 50 aligned nucleotides and reads with more than 2% of ambiguities, or 2% of homopolymers, respectively, were excluded from further processing. Putative contaminations and artefacts read with a low alignment quality (50 alignment identity, 40 alignment score reported by SINA), were identified and excluded from downstream analysis. After these initial steps of quality control, identical reads were identified (dereplication), the unique reads were clustered (OTUs), on a per sample basis, and the reference read of each OTU was classified. Dereplication and clustering were done using cd-hit-est (version 3.1.2; <http://www.bioinformatics.org/cd-hit>) (Li and Godzik, 2006) running in accurate mode, ignoring overhangs, and applying identity criteria of 1.00 and 0.98, respectively. The classification was performed by a local nucleotide BLAST search against the non-redundant version of the SILVA SSU Ref dataset (release 132; <http://www.arb-silva.de>) using blastn (version 2.2.30+; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) with standard settings (Camacho et al., 2009). The classification of each OTU reference read was mapped onto all reads that were assigned to the respective OTU. This yields quantitative information (number of individual reads per taxonomic path), within the

limitations of PCR and sequencing technique biases, as well as, multiple rRNA operons. Reads without any BLAST hits or reads with weak BLAST hits, where the function “($\% \text{ sequence identity} + \% \text{ alignment coverage}$)/2” did not exceed the value of 93, remain unclassified. These reads were assigned to the meta-group “No Relative” in the SILVAngs fingerprint and Krona charts (Ondov et al., 2011). The summary of the reads classified is provided in Table S3.

Table S2: Summary of *P. membranacea* 454 reads.

Tissue	Sequence file id.	Number of reads	Avg.length	Number of reads after trim	Avg.length after trim	Total predicted OTUs (97% similarity)
Apothecia	FV5KH4Y01 (single)	544,701	418	544,687	414.5	195,668
	FV5KH4Y02 (single)	488,366	397.3	488,353	392.9	
	FWE1MRS01 (single)	612,730	405.9	612,715	400.7	
	FWE1MRS02 (single)	594,529	387.2	594,512	381.8	
Thallus	FQ16HFJ01 (single)	597,910	370.3	597,865	366.9	144,943
	FQ16HFJ02 (single)	557,591	325.9	557,543	322.5	
	FQUWX3301 (single)	664,358	375.3	664,340	371.4	
	FQUWX3302 (single)	630,714	346.4	630,687	342.8	

Table S3: Summary of classified reads.

i) Project Name: <i>P. membranacea</i> apothecia LSU	ii) Project Name: <i>P. membranacea</i> thallus SSU
Number of Sequences: 195668	Number of Sequences: 144943
Number of Rejected Sequences: 184261 (94.17 %)	Number of Rejected Sequences: 144696 (99.83 %)
Raw Sequence Information	Raw Sequence Information
Min. Length: 258	Min. Length: 232
Avg. Length: 260	Avg. Length: 233
Max. Length: 260	Max. Length: 233
Aligned Sequence Information	Aligned Sequence Information
Min. Length: 1	Min. Length: 1
Avg. Length: 258	Avg. Length: 228
Max. Length: 260	Max. Length: 233
Quality Information (rejected by)	Quality Information (rejected by)
Alignment BP Score: 1765	Alignment BP Score: 38
Alignment Identity: 159076	Alignment Identity: 134745
Alignment Score: 10230	Alignment Score: 2577
Ambiguous Bases: -	Ambiguous Bases: -
Homopolymers: 11712	Homopolymers: 5763
Quality: -	Quality: -
Length: 1478	Length: 1573
Clustering Information	Clustering Information
Number of OTUs: 11329 (5.79 %)	Number of OTUs: 227 (0.16 %)
Number of Clustered Sequences: 69 (0.04 %)	Number of Clustered Sequences: 15 (0.01 %)
Number of Replicates: 9 (0.00 %)	Number of Replicates: 5 (0.00 %)
Classification Information	Classification Information
Number of Classified Sequences: 352 (0.18 %)	Number of Classified Sequences: 173 (0.12 %)
Number of "No Relative": 11055 (5.65 %)	Number of "No Relative": 74 (0.05 %)
Alignment	Alignment
SINA Version: SINA v1.2.10 for ARB SVN (revision21008)	SINA Version: SINA v1.2.10 for ARB SVN (revision 21008)
Min Align. Identity (%): 40	Min Align. Identity (%): 50
Min Align. Score: 30	Min Align. Score: 40
Min Basepair Score (%): 30	Min Basepair Score (%): 30
Quality Control	Quality Control
Min. Sequence Quality (%): 30	Min. Sequence Quality (%): 30
Min. Length (aligned nuc.): 50	Min. Length (aligned nuc.): 50
Max. Ambiguities (%): 2	Max. Ambiguities (%): 2
Max. Homopolymers (%): 2	Max. Homopolymers (%): 2
Clustering	Clustering
CD-Hit Version: 3.1.2	CD-Hit Version: 3.1.2
Min. OTU Identity (%): 98	Min. OTU Identity (%): 98
Classification	Classification
BLAST Version: 2.2.30+	BLAST Version: 2.2.30+
Reference: SILVA	Reference: SILVA
Reference Version: 132	Reference Version: 132
Similarity (%): 93	Similarity (%): 93

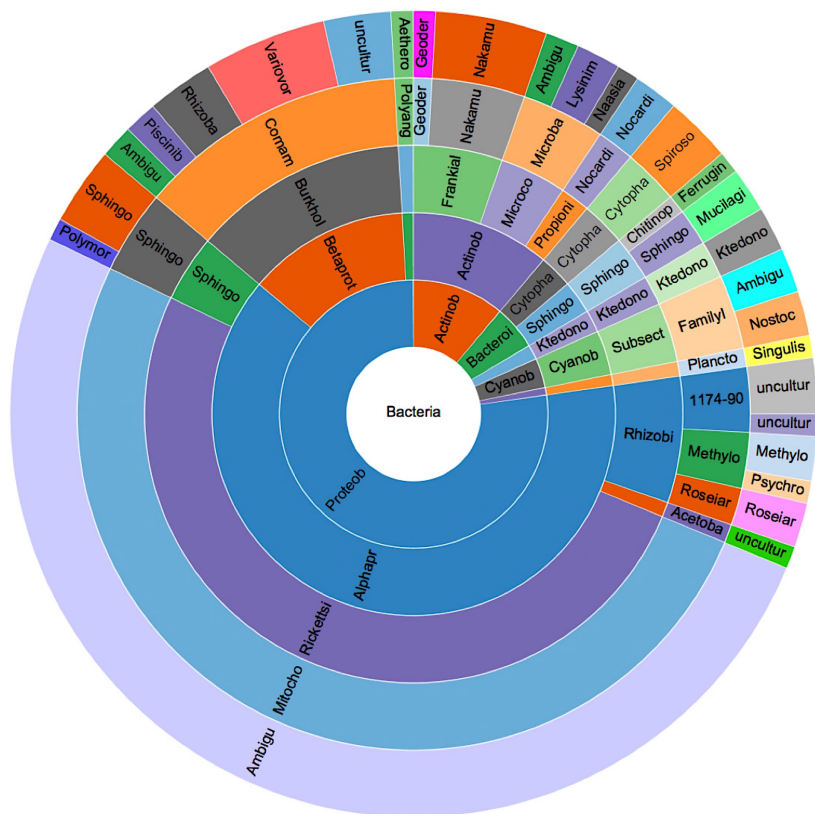




Figure S1: Relative abundance of OTUs of bacterial communities' present in the apothecia of *Peltigera membranacea*.

k-kingdom, p-phylum, c-class, o-order, f-family, g-genus.



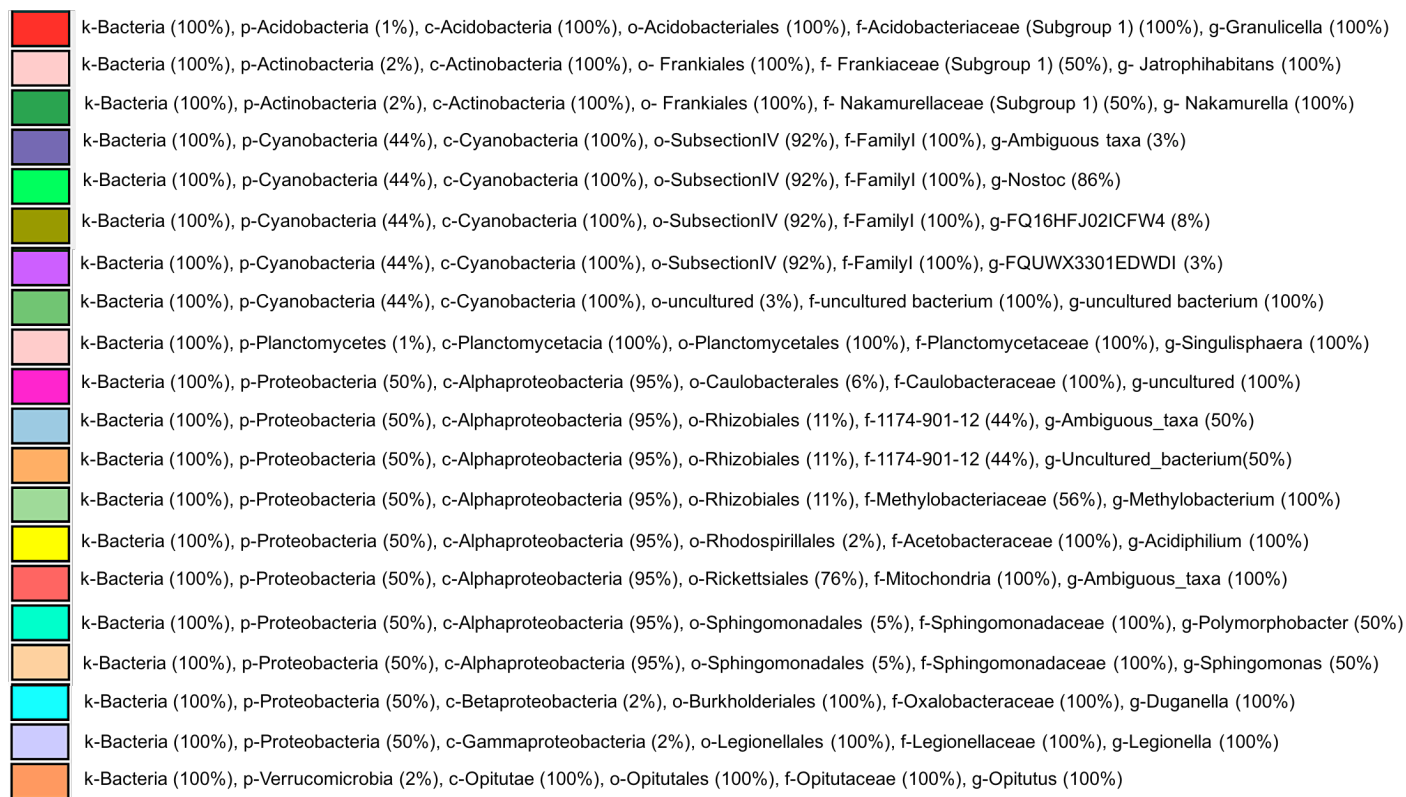
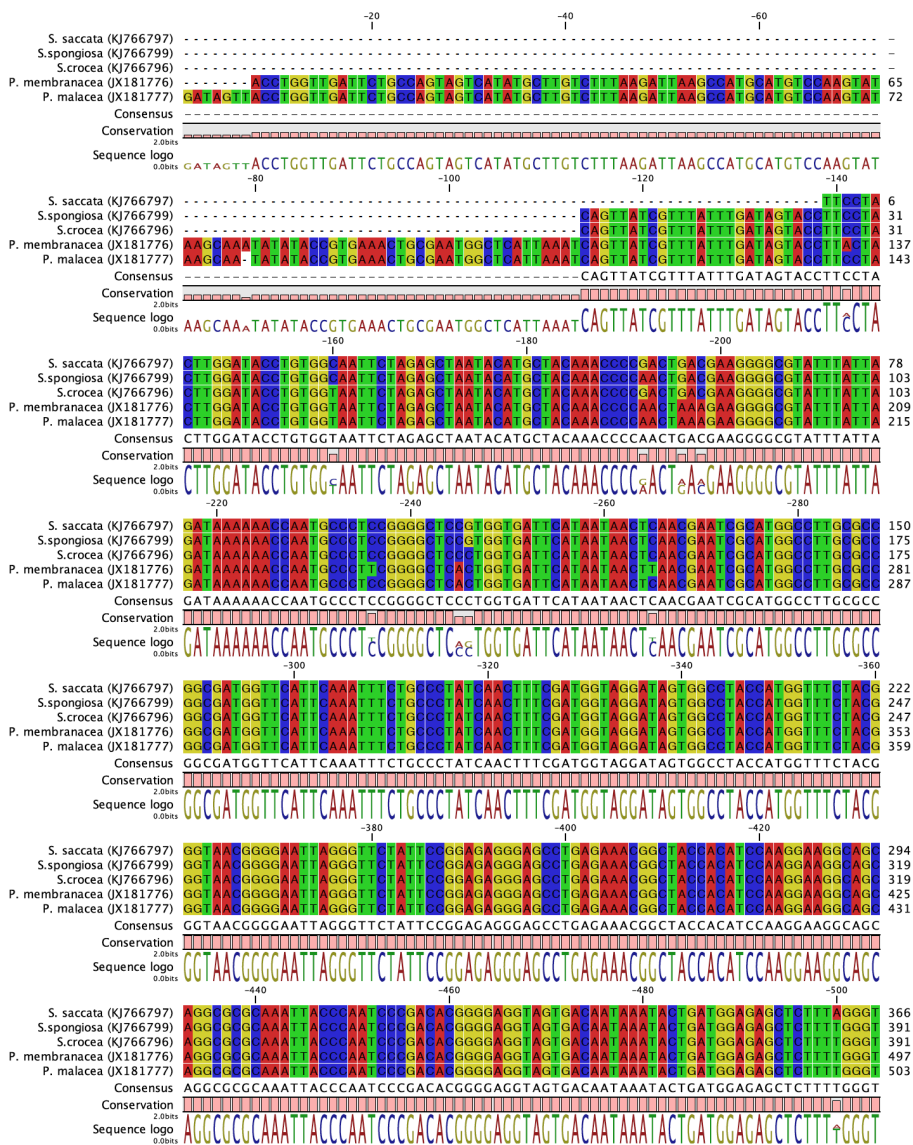


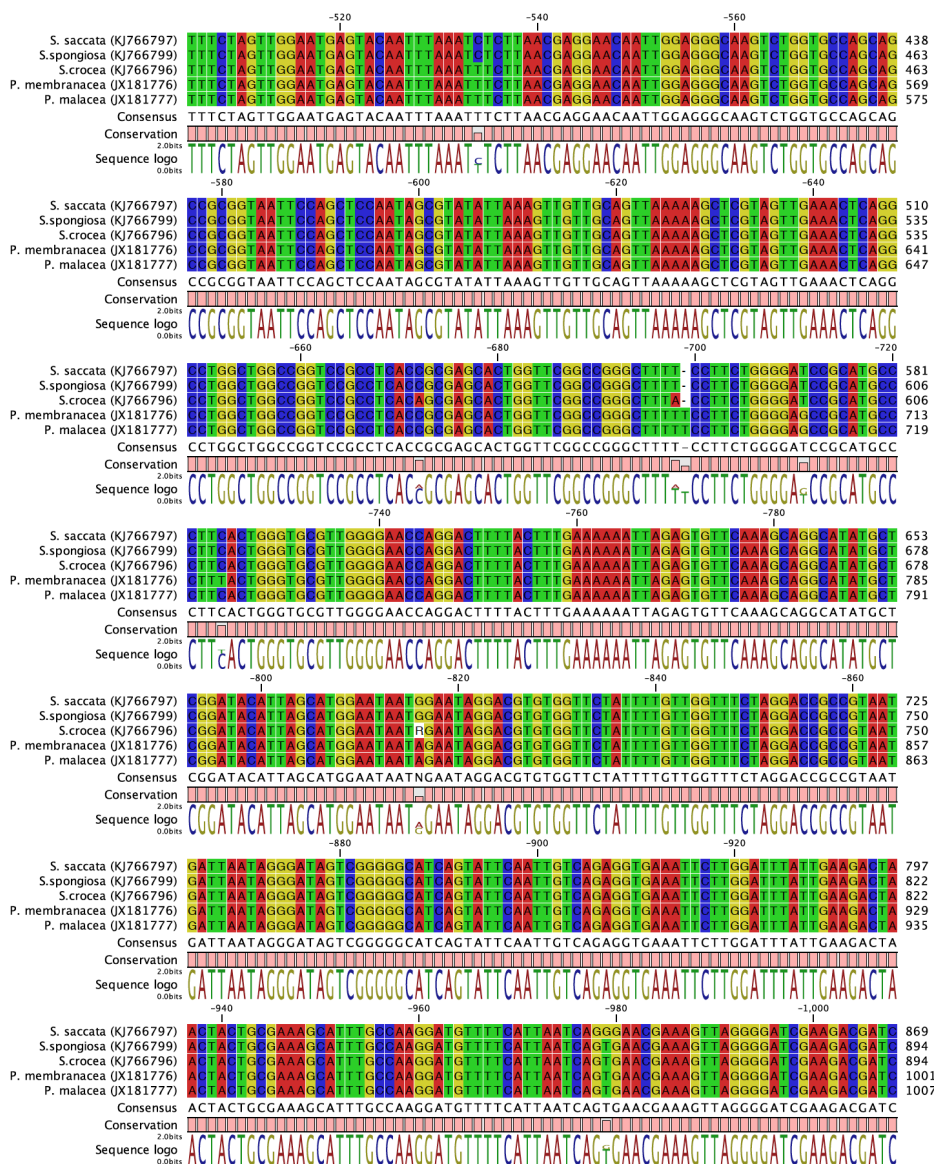
Figure S2: Relative abundance of OTUs of bacterial communities present in the thallus of *Peltigera membranacea*.
k-kingdom, p-phylum, c-class, o-order, f-family, g-genus.

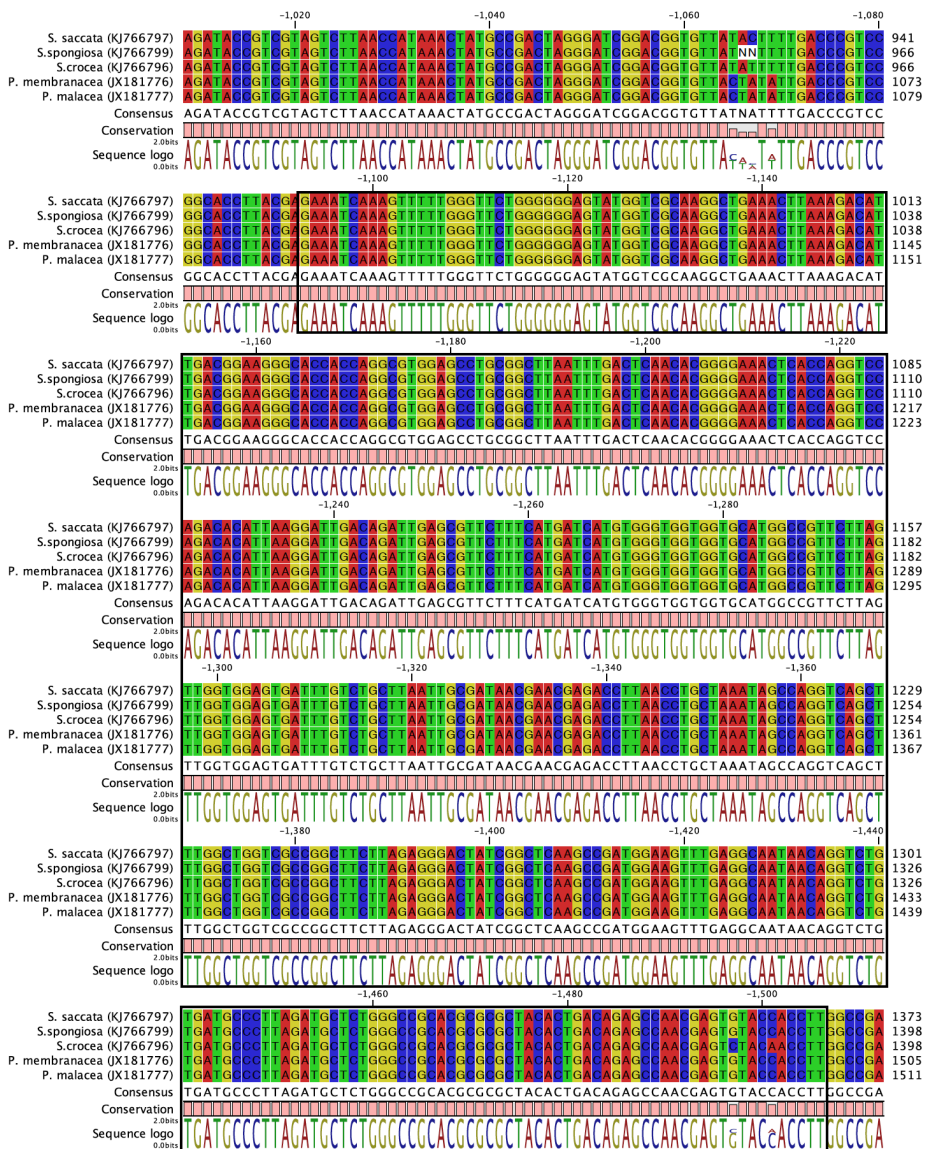
2. Analysis to confirm the presence/absence of more than one fungus (*Solorina* sps.).

For this analysis *P. membranacea* 454 reads from apothecia were mapped to 18S rDNA subunit of *P. membranacea* (accession no: JX181777, 1,803 bp) and *Solorina saccata* (accession no: KJ766797, 1,612 bp) using stringent parameters (match score = 1, mismatch cost = 1, insertion cost = 1, deletion cost = 1, length fraction = 1.0, similarity fraction = 1.0, global alignment = no) using CLC Genomics Workbench v9.5.3. A total of 376 and 42 reads mapped to 18S rDNA of *P. membranacea* and *S. saccata*, respectively. Most of the reads mapped to a 415 bp (942..764 bp) region of *S. saccata* 18S rDNA (Figure S4.B), and therefore the sequence conservation of that region was checked by performing a multiple alignments of the 18S rDNA sequences from *P. membranacea*, *P. malacea* (accession no: JX181777), *S. saccata*, *S. crocea* (accession no: KJ766796), and *S. spongiosa* (accession no: KJ766799) were performed (Figure S3). The 415 bp region was conserved for all the 18S rDNA sequences suggesting that the reads mapped were not specific to *Solorina*. Also, the nucleotide (T) at 1512 position was unique for *P. membranacea*, and the reads mapped at that position (Figure S4.A), whereas a SNP (C) was identified at that position of *Solorina* 18S rDNA region (Figure S4.B).

232







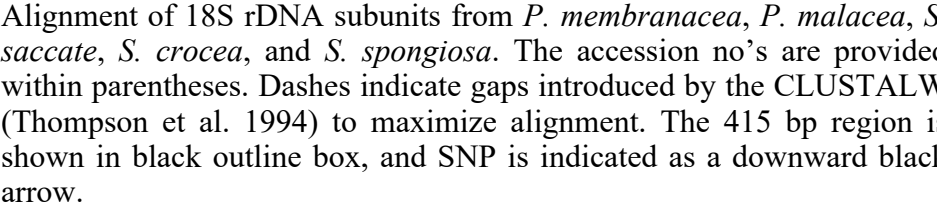
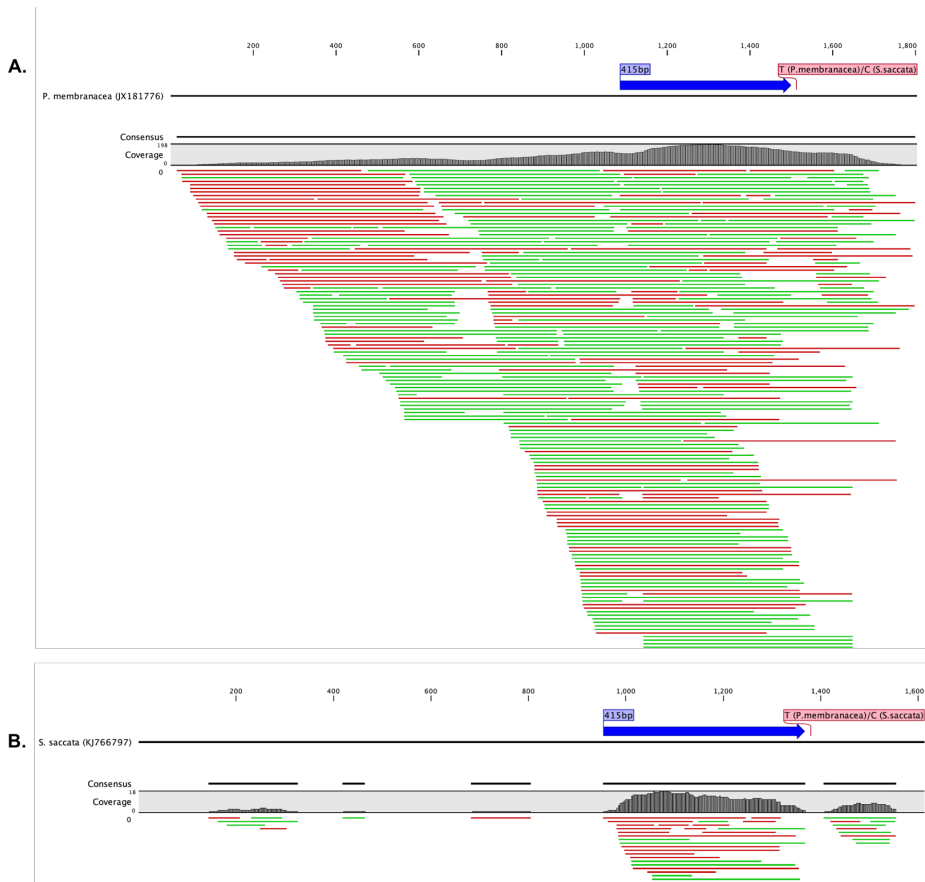
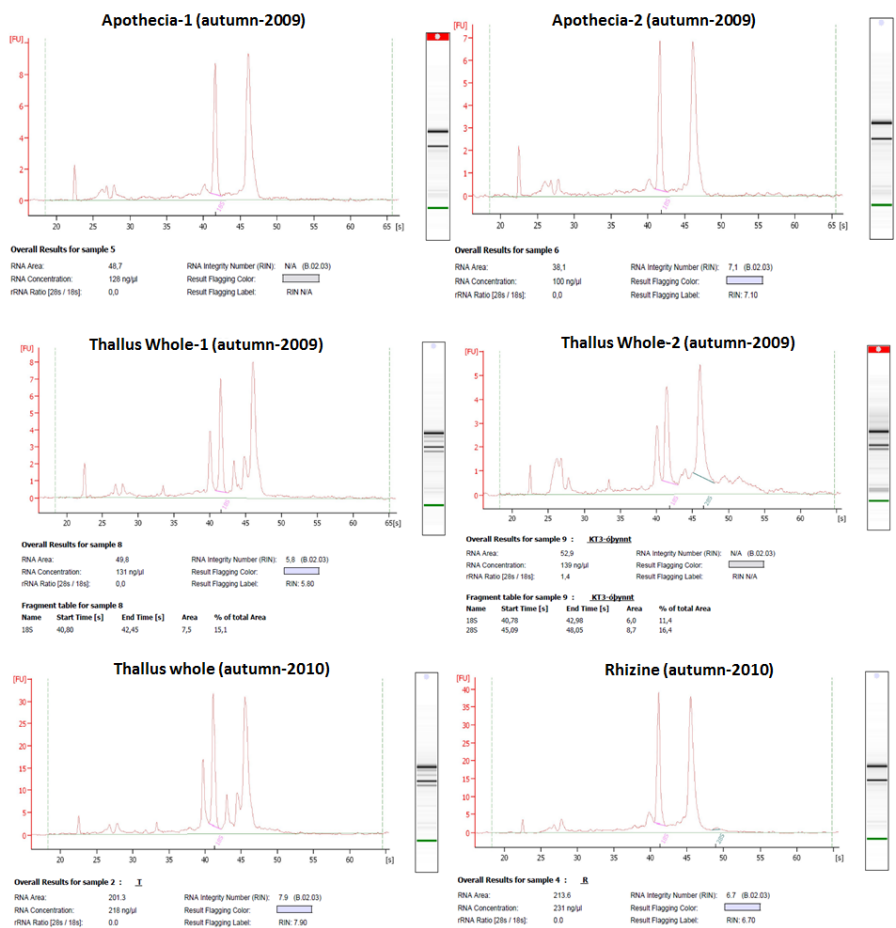


Figure S4: Mapping of 454 apothecial reads to 18S rDNA subunit.

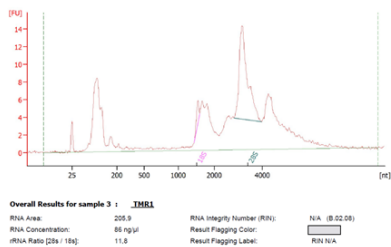


Mapping of 454 apothecial reads to 18S rDNA subunit of (A) *P. membranacea* (accession no: JX181777) and (B) *Solorina saccata* (accession no: KJ766797). The 415 bp conserved region is shown as blue flag and SNP as red flag. The coverage is depicted as grey bar plots. Forward and reverse reads are shown in green and red colors, respectively.

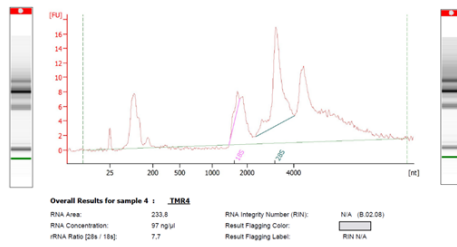
Figure S5: RNA analysis of *P. membranacea* RNA on the Agilent Bioanalyzer 2100.



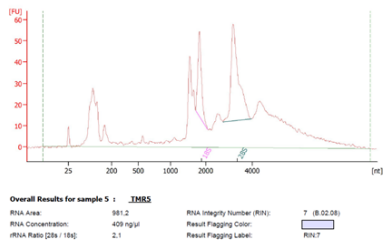
Thallus Base-1 (summer-2012)



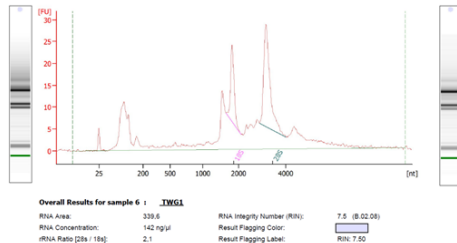
Thallus Base-2 (summer-2012)



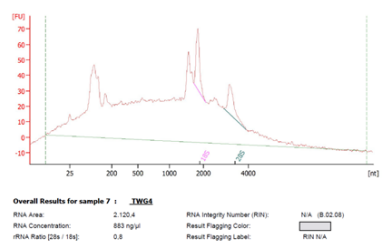
Thallus Base-3 (summer-2012)



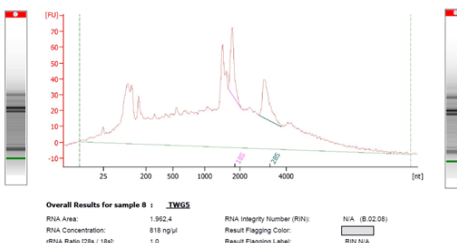
Thallus Whole (G1) (summer-2012)



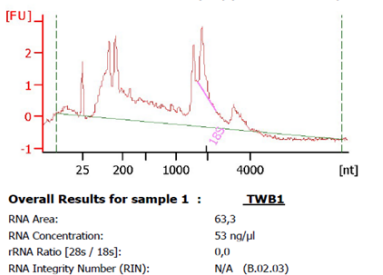
Thallus Whole (G2) (summer-2012)



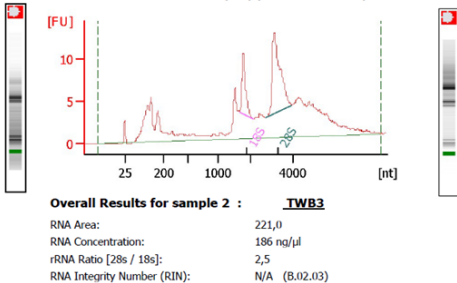
Thallus Whole (G3) (summer-2012)



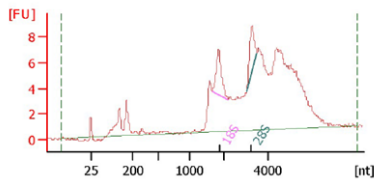
Thallus Whole (B1) (summer-2012)



Thallus Whole (B2) (summer-2012)

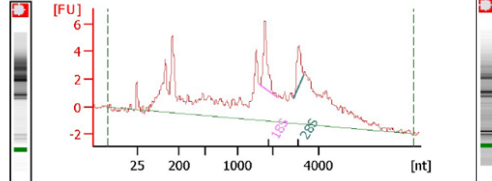


Thallus Whole (R1) (summer-2012)



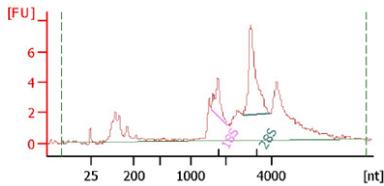
Overall Results for sample 3 : **TWR2**
 RNA Area: 167,7
 RNA Concentration: 141 ng/μl
 rRNA Ratio [28s / 18s]: 0,8
 RNA Integrity Number (RIN): N/A (B.02.03)

Thallus Whole (R2) (summer-2012)



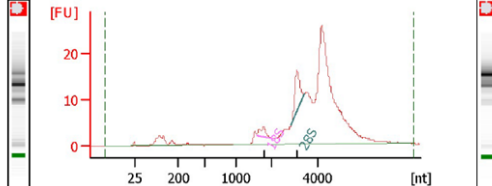
Overall Results for sample 4 : **TWR3**
 RNA Area: 156,7
 RNA Concentration: 132 ng/μl
 rRNA Ratio [28s / 18s]: 0,5
 RNA Integrity Number (RIN): N/A (B.02.03)

Thallus Whole (W1) (summer-2012)



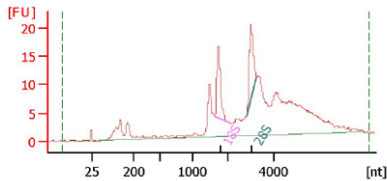
Overall Results for sample 5 : **TWW1**
 RNA Area: 100,4
 RNA Concentration: 85 ng/μl
 rRNA Ratio [28s / 18s]: 2,9
 RNA Integrity Number (RIN): N/A (B.02.03)

Thallus Whole (W2) (summer-2012)



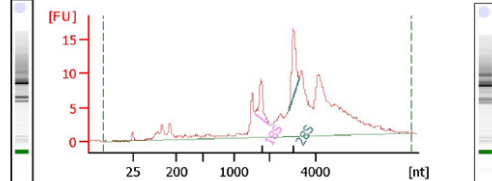
Overall Results for sample 6 : **TWW2**
 RNA Area: 268,9
 RNA Concentration: 227 ng/μl
 rRNA Ratio [28s / 18s]: 3,2
 RNA Integrity Number (RIN): N/A (B.02.03)

Thallus Whole (Y1) (summer-2012)



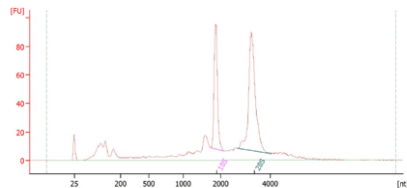
Overall Results for sample 7 : **TWY1**
 RNA Area: 257,7
 RNA Concentration: 217 ng/μl
 rRNA Ratio [28s / 18s]: 1,0
 RNA Integrity Number (RIN): 6,6 (B.02.03)

Thallus Whole (Y2) (summer-2012)



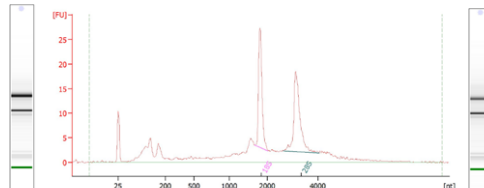
Overall Results for sample 8 : **TWY2**
 RNA Area: 212,1
 RNA Concentration: 179 ng/μl
 rRNA Ratio [28s / 18s]: 1,5
 RNA Integrity Number (RIN): 5,0 (B.02.03)

Rhizine-1 (summer-2012)



Overall Results for sample 1 : **B1**
 RNA Area: 698,5
 RNA Concentration: 291 ng/μl
 rRNA Ratio [28s / 18s]: 1,7
 RNA Integrity Number (RIN): 8,3 (B.02.08)
 Result Flipping Color:
 Result Flipping Label: RIN: 8.30

Rhizine-2 (summer-2012)

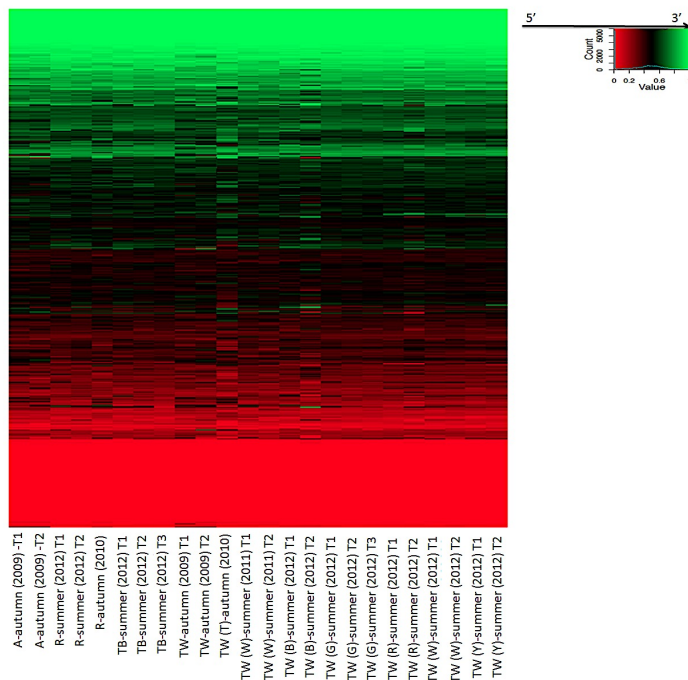


Overall Results for sample 2 : **B2**
 RNA Area: 208,9
 RNA Concentration: 87 ng/μl
 rRNA Ratio [28s / 18s]: 1,1
 RNA Integrity Number (RIN): 7,8 (B.02.08)
 Result Flipping Color:
 Result Flipping Label: RIN: 7.80

Table S4: Summary of transcriptome assemblies.

	Apothecia (au-2009)	Rhizines (au-2010)	Thallus Whole (au-2009)	Thallus Whole (su-2011)
# of Transcripts:	18,551	48,708	10,392	32,441
Total length (nt):	9,880,044	40,912,208	4,986,206	31,867,970
Longest sequence (nt):	7,895	17,356	7,919	18,258
Shortest sequence (nt):	201	201	201	201
Mean sequence length (nt):	533	840	480	982
Median sequence length (nt):	350	341	321	516
N50 sequence length (nt):	668	1,887	558	1,788
L50 sequence length (nt):	3,947	6,327	2,295	5,483
# of sequences > 1K (nt):	2,088	11,424	902	10,537
# of sequences > 5K (nt):	13	668	4	294
# of sequences > 10K (nt):	0	18	0	12

Figure S6: Heat map showing 3' bias



3' bias is calculated as the ratio of reads mapped to the 3' half of the transcript compared to the total length of the transcript. Rows correspond to transcripts and columns to samples. Green indicates a high ratio on the 3' half of the contig and therefore 3' bias.

Appendix C: BUSCO analysis.

Table S5: BUSCO analysis.

	<i>Peltigera membranacea</i>					Reference Genomes					
	Genome	Transcriptomes									
	Pmem v1.1	Apothecia (au-2009)	Rhizine (au-2009)	Thallus Whole (au-2009)	Thallus Whole (su-2011)	<i>Aspergillus nidulans</i>	<i>Diplocarpon rosae*</i>	<i>Marssonina brunnea*</i>	<i>Botrytis cinerea*</i>	<i>Sclerotinia sclerotiorum*</i>	<i>Magnaporthea oryzae*</i>
Complete BUSCOs	1295	420	1356	187	1219	1422	1385	1412	1423	1387	1418
Complete + fragmented BUSCOs	1394	1078	1419	592	1407	1433	1426	1432	1435	1429	1428
Complete and single-copy BUSCOs	1285	417	1011	186	909	1415	277	1337	1340	1301	1338
Complete and duplicated BUSCOs	10	3	345	1	310	7	1108	75	83	86	80
Fragmented BUSCOs	99	658	63	405	188	11	41	20	12	42	10
Missing BUSCOs	44	360	19	846	31	5	12	6	3	9	10
Total BUSCO groups searched	1438	1438	1438	1438	1438	1438	1438	1438	1438	1438	1438

* Adapted from "A draft genome sequence of the rose black spot fungus *Diplocarpon rosae* reveals a high degree of genome duplication", by Neu E, Featherston J, Rees J, Debener T (2017) PLoS ONE 12(10): e0185310. <https://doi.org/10.1371/journal.pone.0185310>

Table S6: Analysis of missing BUSCO proteins from *P. membranacea* genome (P.mem v1.1).

S.No	Missing BUSCOs	Number of HSPs	Lowest E-value	Hits	Greatest identity %	Greatest positive %
1	EOG708W9T	3	3.44E-36	scaff36184	79.2	90.9
2	EOG70KH1C	2	3.44E-34	scaff31458	64.8	82.4
3	EOG71CFZ8	1	0.0652955	scaff31940	40.7	74.1
4	EOG71K6FR	2	1.19E-45	scaff34475	55.9	71.7
5	EOG741ZC1	2	3.96515	scaff34962	47.5	57.5
6	EOG757D9B	2	6.47E-14	scaff38795	77.3	90.9
7	EOG7712MK	4	7.35E-13	scaff30836	76.5	90.2
8	EOG77WWR0	3	1.24E-46	scaff30869	57.1	67.9
9	EOG780RWX	4	6.79E-15	scaff32887	81.8	90.9
10	EOG785QVJ	2	2.84539	scaff38123	36.8	54.0
11	EOG7966TJ	0		not available		
12	EOG7BP8B9	1	2.13E-10	scaff32461	28.2	41.8
13	EOG7CNZTF	2	4.56E-40	scaff33611	50.3	67.6
14	EOG7CRTZN	0		not available		
15	EOG7DJSXH	24	5.50E-24	scaff33737	62.7	83.1
16	EOG7DZ8Z5	3	1.25E-15	scaff36119	75.0	80.0
17	EOG7FFN53	1	5.73E-05	scaff35230	81.0	100.0
18	EOG7FR7TW	4	2.60E-10	scaff40134	70.6	92.2
19	EOG7FV40Z	5	3.92E-46	scaff31417	61.1	88.9
20	EOG7FZ09D	5	1.02E-22	scaff34526	78.4	94.1
21	EOG7GXPR2	6	1.60E-47	scaff31570	90.0	95.0
22	EOG7HHX4K	3	7.80E-13	scaff31417	76.9	89.7
23	EOG7HDX37	4	9.54E-09	scaff52152	85.0	90.0
24	EOG7J188V	1	3.28E-08	scaff30664	36.9	52.4
25	EOG7K0ZQF	5	6.63E-46	scaff34409	84.8	92.4
26	EOG7K3V05	3	1.82E-15	scaff32960	41.4	54.3
27	EOG7KDFMZ	6	1.96E-04	scaff33607	42.9	53.9
28	EOG7KDFPN	1	0.041606	scaff34802	67.6	78.4
29	EOG7KM654	4	5.27E-16	scaff34890	86.5	89.2
30	EOG7KQ2CS	1	5.97242	scaff32252	39.4	57.6
31	EOG7KWSVP	3	9.48E-19	scaff33998	64.9	86.0
32	EOG7N0CHC	8	2.88E-43	scaff31343	98.7	100.0
33	EOG7QNVZZ	0		not available		
34	EOG7QRR76	2	3.40E-12	scaff36184	100.0	100.0
35	EOG7QVMD6	7	1.58E-35	scaff30687	89.3	90.9
36	EOG7SJDGJ	5	6.73E-03	scaff32455	85.0	85.0
37	EOG7TJ3TQ	76	4.75E-63	scaff35032	71.2	79.7
38	EOG7TQVCM	6	1.85E-23	scaff51413	81.3	93.8
39	EOG7VQJQ2	1	1.51E-03	scaff45396	32.0	46.7
40	EOG7WDNDF	2	4.60E-56	scaff35501	52.6	64.9
41	EOG7X0VRZ	0		not available		
42	EOG7XH72F	22	2.71E-16	scaff37663	70.6	88.2
43	EOG7XST50	4	1.12E-21	scaff36406	68.3	84.1
44	EOG7ZPNX4	1	5.60832	scaff31197	58.8	82.4

Appendix D: CEGMA Analysis.

The P.mem v1.1 assembly was found to contain complete (full-length) copies of 233 of the 248 CEGs, with an additional nine CEGs being partially represented and six CEGs missing (Table 2.3). The six missing CEGs from *P. membranacea* genome were further investigated, and (i) three CEG's, KOG0969, KOG1775, and KOG3318 were missing only from the *P. membranacea* genome; (ii) two CEG's, KOG0346, and KOG0741 were missing from both *P. membranacea* and *A. nidulans*; and (iii) KOG0894 was missing from all three (Tables S7, S8 and S9). The *P. membranacea* transcriptome assembly was found to contain complete copies of 234 CEGs, with six CEGs being partially represented and eight CEGs missing (Table S8). Out of the eight CEGs missing from the *P. membranacea* transcriptome, seven CEGs were represented in the *P. membranacea* genome assembly with KOG0984 missing from both. For comparison, a well assembled *Aspergillus nidulans* genome (Cerqueira et al. 2014) was analyzed with this method, and the genome contained complete copies of 239 CEGs, with two partially represented CEGs and seven CEGs missing (Table S9). A local TBLASTN was performed (Altschul et al. 1990) for the one missing KOG0894 (mitogen-activated protein kinase (354aa) CEG. Accession: XP_962163.2 from *Neurospora crassa* was used as a query and was searched against the P.mem v1.1 genome and the Trinity transcriptomes. KOG0984 was found to be present in the genome [scaffold 32061 (length=57,422; evalue 5.96E-170)], apothecia transcript [767 (length 2721; evalue 0.0)], thallus transcript [1300 (length=2690; evalue 0.0)] as well as in rhizine transcript [643 (length 2465; evalue 0.0)], and the AUGUSTUS (see section 2.3.3) derived gene [g4890 (length=356aa, evalue 0.0)]. A mitogen-activated protein kinase was also identified in *Aspergillus nidulans*, accession no: AAF12815.

Table S7: CEGMA report on *P. membranacea* genome assembly P.mem v1.1).

# Statistics of the completeness of the P. membranacea genome based on 248 CEGs #						
	#Prots	%Completeness	-	#Total	Average	%Ortho
Complete	233	93.95	-	286	1.23	19.31
Group 1	60	90.91	-	71	1.18	16.67
Group 2	52	92.86	-	65	1.25	23.08
Group 3	59	96.72	-	76	1.29	20.34
Group 4	62	95.38	-	74	1.19	17.74
Partial	239	96.37	-	306	1.28	23.43
Group 1	62	93.94	-	74	1.19	17.74
Group 2	53	94.64	-	67	1.26	24.53
Group 3	61	100.00	-	86	1.41	29.51
Group 4	63	96.92	-	79	1.25	22.22
#	These results are based on the set of genes selected by Genis Parra					#
#	Key:					#
#	Prots = number of 248 ultra-conserved CEGs present in genome					#
#	%Completeness = percentage of 248 ultra-conserved CEGs present					#
#	Total = total number of CEGs present including putative orthologs					#
#	Average = average number of orthologs per CEG					#
#	%Ortho = percentage of detected CEGs that have more than 1 ortholog					#
##	Missing proteins			##		
# Complete						
	KOG0181, KOG0346, KOG0650, KOG0741, KOG0894, KOG0969, KOG1185, KOG1466, KOG1775, KOG2311, KOG2531, KOG2906, KOG3285, KOG3318, KOG3499					
# Partial						
	KOG0181, KOG0650, KOG1185, KOG1466, KOG2311, KOG2531, KOG2906, KOG3285, KOG3499					

Table S8: CEGMA report on *P. membranacea* Trinity transcriptome assembly.

# Statistics of the completeness of the P. membranacea transcriptome based on 248 CEGs #						
	#Prots	%Completeness	-	#Total	Average	%Ortho
Complete	234	94.35	-	633	2.71	76.50
Group 1	59	89.39	-	154	2.61	74.58
Group 2	53	94.64	-	132	2.49	73.58
Group 3	57	93.44	-	145	2.54	68.42
Group 4	65	100.00	-	202	3.11	87.69
Partial	242	97.58	-	781	3.23	88.02
Group 1	61	92.42	-	176	2.89	81.97
Group 2	56	100.00	-	177	3.16	87.50
Group 3	60	98.36	-	193	3.22	85.00
Group 4	65	100.00	-	235	3.62	96.92
#	These results are based on the set of genes selected by Genis Parra					#
#	Key:					#
#	Prots = number of 248 ultra-conserved CEGs present in genome					#
#	%Completeness = percentage of 248 ultra-conserved CEGs present					#
#	Total = total number of CEGs present including putative orthologs					#
#	Average = average number of orthologs per CEG					#
#	%Ortho = percentage of detected CEGs that have more than 1 ortholog					#
##	Missing proteins			##		
#	Complete					
	KOG0025, KOG0188, KOG0477, KOG0894, KOG0948, KOG0964, KOG1088, KOG1185, KOG1466, KOG2004, KOG2531, KOG2785, KOG3013, KOG3285					
#	Partial					
	KOG0188, KOG0964, KOG1185, KOG1466, KOG2531, KOG2785					

Table S9: CEGMA report on *Aspergillus nidulans* genome.

#	Statistics of the completeness of the A.nidulans genome based on 248 CEGs						#
	#Prots	%Completeness	-	#Total	Average	%Ortho	
Complete	239	96.37	-	267	1.12	8.37	
Group 1	64	96.97	-	71	1.11	6.25	
Group 2	52	92.86	-	55	1.06	5.77	
Group 3	60	98.36	-	72	1.20	15.00	
Group 4	63	96.92	-	69	1.10	6.35	
Partial	246	99.19	-	280	1.14	10.57	
Group 1	66	100.00	-	73	1.11	6.06	
Group 2	54	96.43	-	59	1.09	9.26	
Group 3	61	100.00	-	74	1.21	16.39	
Group 4	65	100.00	-	74	1.14	10.77	
#	These results are based on the set of genes selected by Genis Parra						#
#	Key:						#
#	Prots = number of 248 ultra-conserved CEGs present in genome						#
#	%Completeness = percentage of 248 ultra-conserved CEGs present						#
#	Total = total number of CEGs present including putative orthologs						#
#	Average = average number of orthologs per CEG						#
#	%Ortho = percentage of detected CEGs that have more than 1 ortholog						#
##	Missing proteins				##		
# Complete							
	KOG0261, KOG0346, KOG0462, KOG0741, KOG0894, KOG1088, KOG1637, KOG1795, KOG2311						
# Partial							
	KOG0462, KOG2311						

Table S10: Presence/absence of missing CEGMA proteins from *P. membranacea* genome (P.mem v1.1), *P. membranacea* Trinity transcriptome and *A. nidulans* genome.

Missing CEGs	<i>P. membranacea</i> genome (P.mem v1.1)	<i>P. membranacea</i> Trinity transcriptome	<i>A. nidulans</i> genome
KOG0025	+	-	+
KOG0188	+	-	+
KOG0261	+	+	-
KOG0346	-	+	-
KOG0477	+	-	+
KOG0741	-	+	-
<u>KOG0894</u>	-	-	-
KOG0948	+	-	+
KOG0969	-	+	+
KOG1088	+	+	-
KOG1637	+	+	-
KOG1775	-	+	+
KOG1795	+	+	-
KOG2004	+	-	+
KOG3013	+	-	+
KOG3318	-	+	+

As of 2015, CEGMA has been effectively replaced by BUSCO, as CEGMA is no longer supported and actively updated.

Appendix E: Comparison of the *de novo* transcriptome assembly and reference genome annotation.

Table S11: List of missing, and unmapped *de novo* Trinity transcripts in P.mem v1.1 genome along with BLAST hits.

<i>De novo</i> transcriptomes	No of Transcripts missing in P.mem v1.1	No of Transcripts mapped to P.mem v1.1	No of Transcripts unmapped to P.mem v1.1	Blast Hits	Hypothetical proteins
Rhizines (au-2009)	330	79	251	73	26
Apothecia (au-2009)	1623	1349	274	115	25
Thallus Whole (au-2009)	9	7	2	0	0
Thallus Whole (su-2011)	106	72	34	13	7

Table S12: BLAST results for the unmapped Rhizine (au-2009) Trinity *de novo* transcripts

SeqName	Description
comp44222_c0_seq1	ADP ATP translocase 2
comp18766_c0_seq1	hypothetical protein M408DRAFT_21368
comp19535_c0_seq1	60S ribosomal L12
comp18052_c0_seq1	histone H2A
comp40277_c0_seq1	pectin lyase
comp40602_c0_seq1	copper zinc superoxide dismutase
comp41099_c0_seq1	glutathione transferase omega-1
comp40448_c0_seq1	probable 40S ribosomal S12
comp40347_c0_seq1	Dynein heavy chain-2 domain-containing
comp27142_c0_seq1	Pheromone B alpha 2 receptor
comp15934_c0_seq1	ATP-dependent RNA helicase dhh1
comp15545_c0_seq1	hypothetical protein GYMLUDRAFT_253045
comp4633_c0_seq1	hypothetical protein ASPACDRAFT_56193
comp35005_c0_seq1	hypothetical protein M427DRAFT_95618, partial
comp35814_c0_seq1	alpha subunit
comp35807_c0_seq1	hypothetical protein PTTG_31188
comp35360_c0_seq1	hypothetical protein PV08_06397
comp35984_c0_seq1	related to secreted -Streptomyces sviveus

comp34204_c0_seq1	probable Catalase-peroxidase 2
comp35860_c0_seq1	hypothetical protein M408DRAFT_197827
comp9369_c0_seq2	hypothetical protein LY90DRAFT_76715
comp8867_c0_seq1	hypothetical protein LRAMOSA04081
comp8426_c0_seq5	hypothetical protein M422DRAFT_60040
comp13331_c0_seq1	hypothetical protein BCR41DRAFT_343865
comp12676_c0_seq4	related to alpha glucosidase II beta subunit
comp13096_c0_seq1	aryl-alcohol oxidase
comp13096_c0_seq2	aryl-alcohol oxidase
comp16374_c0_seq1	NADH dehydrogenase subunit 1 (mitochondrion)
comp42795_c0_seq1	cell division control 54
comp42777_c0_seq1	G- gamma subunit
comp42937_c0_seq1	cytochrome c oxidase subunit 3 (mitochondrion)
comp43683_c0_seq1	DChain Molecular Architecture Of The Rotary Motor In Atp Synthase From Yeast Mitochondria
comp42296_c0_seq1	Ribosomal L13e
comp37108_c0_seq1	hypothetical protein PIIN_10350
comp36031_c0_seq1	copper radical oxidase
comp37293_c0_seq1	hypothetical protein M408DRAFT_14833
comp36497_c0_seq1	60S ribosomal L3
comp890_c0_seq6	hypothetical protein ASPACDRAFT_37804
comp32718_c0_seq1	carbohydrate esterase family 4
comp33544_c0_seq1	related to EXG1-exo-beta-1,3-glucanase (I II) major isoform
comp33717_c0_seq1	hypothetical protein MVEG_12081
comp33082_c0_seq1	hypothetical protein BOTBODRAFT_167254
comp33200_c0_seq1	hypothetical protein HMPREF1541_09567
comp21454_c0_seq2	Glutamate mitochondrial
comp21012_c0_seq1	hypothetical protein HMPREF1541_03070
comp20086_c0_seq1	hypothetical protein V502_03873
comp22600_c0_seq1	stearoyl- desaturase (delta-9 desaturase)
comp22038_c0_seq1	hypothetical protein BDEG_23708
comp22813_c0_seq1	hypothetical protein AMAG_13861
comp25419_c0_seq1	glycoside hydrolase family 44
comp24608_c0_seq1	water dikinase
comp25691_c0_seq2	60S ribosomal L7
comp39654_c0_seq1	acyl- dehydrogenase domain-containing
comp39082_c0_seq1	nucleoside phosphatase GDA1

comp38310_c0_seq1	glycoside hydrolase family 47
comp38499_c0_seq1	DUF1900-domain-containing protein
comp39640_c0_seq1	hypothetical protein
comp39650_c0_seq1	tyrosine phosphatase Pyp2
comp39155_c0_seq1	HSP70-domain-containing protein
comp10130_c0_seq1	hypothetical protein V502_03873
comp10084_c0_seq2	peptidase C1A
comp10820_c0_seq1	cytochrome oxidase subunit 3 (mitochondrion)
comp10672_c0_seq1	YLR154W-A-like protein
comp10095_c0_seq2	lectin
comp7349_c0_seq5	serine-type endopeptidase
comp6279_c0_seq1	creatine kinase M-type
comp6279_c0_seq2	creatine kinase M-type
comp28860_c0_seq1	hypothetical protein PIIN_09540
comp28571_c0_seq1	hypothetical protein V502_03873
comp29148_c0_seq1	hypothetical protein K493DRAFT_315500
comp28124_c0_seq1	aminopeptidase
comp30574_c0_seq1	cortical actin cytoskeleton asp1
comp30525_c0_seq1	glycoside hydrolase family 76

Table S13: BLAST results for the unmapped apothecia (au-2009) Trinity *de novo* transcripts

SeqName	Description
comp15674_c0_seq1	ribosomal l30
comp12912_c0_seq1	RNA-directed RNA polymerase
comp17529_c0_seq1	TPR repeat
comp274_c0_seq1	Isoleucine--tRNA ligase
comp11407_c0_seq1	RNA polymerase II transcription factor B subunit 5
comp12777_c0_seq1	hypothetical protein M426DRAFT_28031
comp5262_c0_seq1	glycosyltransferase family 2
comp5785_c0_seq1	C6 finger domain-containing
comp13349_c0_seq1	mRNA splicing factor
comp17063_c0_seq1	potassium channel regulatory factor
comp16974_c0_seq1	transcription factor
comp6802_c0_seq1	RNA-dependent RNA
comp6006_c0_seq1	RNA-dependent RNA
comp15581_c0_seq1	conserved hypothetical protein
comp17067_c0_seq1	hypothetical protein EPUS_01670
comp3250_c0_seq1	translation initiation factor SUI1
comp2993_c0_seq1	di tri peptide transporter 2
comp18026_c0_seq1	undecaprenyl diphosphate synthase
comp8683_c0_seq1	hypothetical protein V496_00252
comp12756_c0_seq1	HC-toxin efflux carrier TOXA
comp12827_c0_seq1	RNA-dependent RNA
comp2897_c0_seq1	Translation machinery-associated 20
comp2704_c0_seq1	PREDICTED: uncharacterized protein LOC105852951
comp5_c0_seq1	hypothetical protein AN5245.2
comp18326_c0_seq1	Zn2 Cys6 DNA-binding
comp4649_c0_seq1	alpha- glucosyltransferase alg-10
comp8677_c0_seq1	Tetratricopeptide-like helical domain
comp8461_c0_seq1	replication-associated poly
comp13889_c0_seq1	uncharacterized protein PAC_02131
comp14630_c0_seq1	MFS general substrate transporter

comp4966_c0_seq1	Something about silencing 4
comp13430_c0_seq1	hypothetical protein LPUS_02750
comp13072_c0_seq1	Pkinase-domain-containing protein
comp13455_c0_seq1	efflux pump antibiotic resistance
comp7099_c0_seq1	glutamate decarboxylase
comp15730_c0_seq1	Pyridoxal phosphate-dependent major subdomain 1
comp15959_c0_seq1	Calcium calmodulin-dependent calcium-dependent kinase
comp5843_c0_seq1	Cytochrome P450
comp4823_c0_seq1	hypothetical protein ASPZODRAFT_132016
comp9608_c0_seq1	hypothetical protein LPUS_07294
comp17766_c0_seq1	DNA mismatch repair PMS1
comp7381_c0_seq1	Trafficking particle complex subunit 2
comp18489_c0_seq1	histone acetyltransferase esa1
comp4010_c0_seq1	peptide transporter ptr2-a
comp4393_c0_seq1	UNKNOWN
comp4519_c0_seq1	hypothetical protein EMCG_04632
comp4120_c0_seq1	acetyl- C-acetyltransferase
comp8106_c0_seq1	methyltransferase protease helicase
comp13340_c0_seq1	replication-associated poly
comp16571_c0_seq1	quinone oxidoreductase
comp2023_c0_seq1	NADH-ubiquinone oxidoreductase kDa subunit
comp14540_c0_seq1	mRNA capping nucleoside-
comp5818_c0_seq1	alpha-1,6-mannosyltransferase subunit
comp13293_c0_seq1	dna mismatch repair pms1
comp1238_c0_seq1	hypothetical protein SLINC_2054
comp10818_c0_seq1	DNA repair RAD5
comp11778_c0_seq1	RNA-dependent RNA polymerase
comp13668_c0_seq1	proline oxidase
comp17090_c0_seq1	hypothetical protein AYL99_09683
comp34_c0_seq1	Immunoglobulin E-set
comp1454_c0_seq1	RNA-dependent RNA
comp10764_c0_seq1	RNA-dependent RNA
comp6975_c0_seq1	replicase

comp14909_c0_seq1	ORF (Fragment)
comp15009_c0_seq1	aspartate aminotransferase
comp11184_c0_seq1	TIM44 subunit of mitochondria import inner membrane translocase
comp11062_c0_seq1	hypothetical protein
comp15431_c0_seq1	hypothetical protein FE78DRAFT_94360
comp15309_c0_seq1	XAP5 protein
comp17019_c0_seq1	beta-lactamase transpeptidase
comp56_c0_seq1	RNA-dependent RNA
comp81_c0_seq1	replicase
comp9563_c0_seq1	hypothetical protein
comp16902_c0_seq1	Zinc RING FYVE PHD-type
comp16061_c0_seq1	mitochondrial import inner membrane translocase subunit TIM44
comp1733_c0_seq1	C2 domain
comp7941_c0_seq1	polyprotein
comp18578_c0_seq1	acyl- thioester hydrolase
comp15096_c0_seq1	DUF962-domain-containing protein
comp11868_c0_seq1	ORF1 poly
comp18680_c0_seq1	afln vera monooxygenase
comp15150_c0_seq1	GPI-anchor biosynthetic
comp16788_c0_seq1	kinase and ribonuclease
comp6506_c0_seq1	NADH alpha subunit 2
comp78_c0_seq1	hypothetical protein BCR34DRAFT_619182
comp10579_c0_seq1	zinc-binding alcohol dehydrogenase
comp7888_c0_seq1	glycosylphosphatidylinositol anchor biosynthesis 11
comp4669_c0_seq1	hypothetical protein FgHV1gp2
comp3852_c0_seq1	RNA dependent RNA polymerase
comp11587_c0_seq1	hypothetical protein SNOG_15762
comp9889_c0_seq1	hypothetical protein CIHG_07200
comp3868_c0_seq1	peptide transport PTR2
comp2816_c0_seq1	glycosyltransferase family 2
comp11172_c0_seq1	Cytochrome P450
comp18381_c0_seq1	deoxyribose-phosphate aldolase

comp1298_c0_seq1	hypothetical protein AOL_s00188g86
comp8554_c0_seq1	Phox-associated domain
comp2945_c0_seq1	replication-associated poly
comp149_c0_seq1	predicted protein
comp1320_c0_seq1	C2 domain
comp6019_c0_seq1	hypothetical protein CC77DRAFT_1030282
comp5351_c0_seq1	replicase
comp11251_c0_seq1	ankyrin repeat domain-containing
comp15147_c0_seq1	hypothetical protein B0A51_14118
comp459_c0_seq1	glycoside hydrolase family 128
comp10232_c0_seq1	conserved hypothetical protein, partial
comp18330_c0_seq1	related to DUF890 domain
comp10291_c0_seq1	coiled-coil domain-containing 25
comp10096_c0_seq1	Solute carrier family 35 member F3
comp14452_c0_seq1	TPA_exp: RNA exonuclease
comp11260_c0_seq1	arginine N-methyltransferase 2
comp14606_c0_seq1	hypothetical protein BAUCODRAFT_142160
comp3211_c0_seq1	transcriptional corepressor
comp10676_c0_seq1	hypothetical protein ASPSYDRAFT_56841
comp403_c0_seq1	hypothetical protein SETTUDRAFT_164344

Table S14:BLAST results for the unmapped *Thallus* whole (au-2009) Trinity *de novo* transcripts.

Seqname	Description
comp6510_c0_seq1	acetyltransferase (GNAT) family domain-containing protein
comp6756_c0_seq1	DUF1771-domain-containing protein

Table S15: BLAST results for the unmapped *Thallus* whole (su-2011) Trinity *de novo* transcripts

SeqName	Description
comp15655_c0_seq1	Cupin conserved barrel domain
comp4449_c0_seq3	hypothetical protein CTHBC1_2972
comp4001_c0_seq4	Uncharacterised protein
comp5025_c0_seq3	extracellular 14-2
comp16511_c0_seq1	hypothetical protein 2
comp17187_c0_seq1	hypothetical protein T4E_5477
comp17509_c0_seq1	hypothetical protein
comp17364_c0_seq1	hypothetical protein
comp16138_c0_seq1	reducing type I polyketide synthase
comp21838_c0_seq1	ribosomal S2
comp20096_c0_seq1	hypothetical protein HMPREF9610_00001, partial
comp23138_c0_seq1	U6 snRNA-associated Sm LSm1
comp7437_c0_seq1	hypothetical protein EW35_3223

Appendix F: Alignment of predicted and curated proteins.

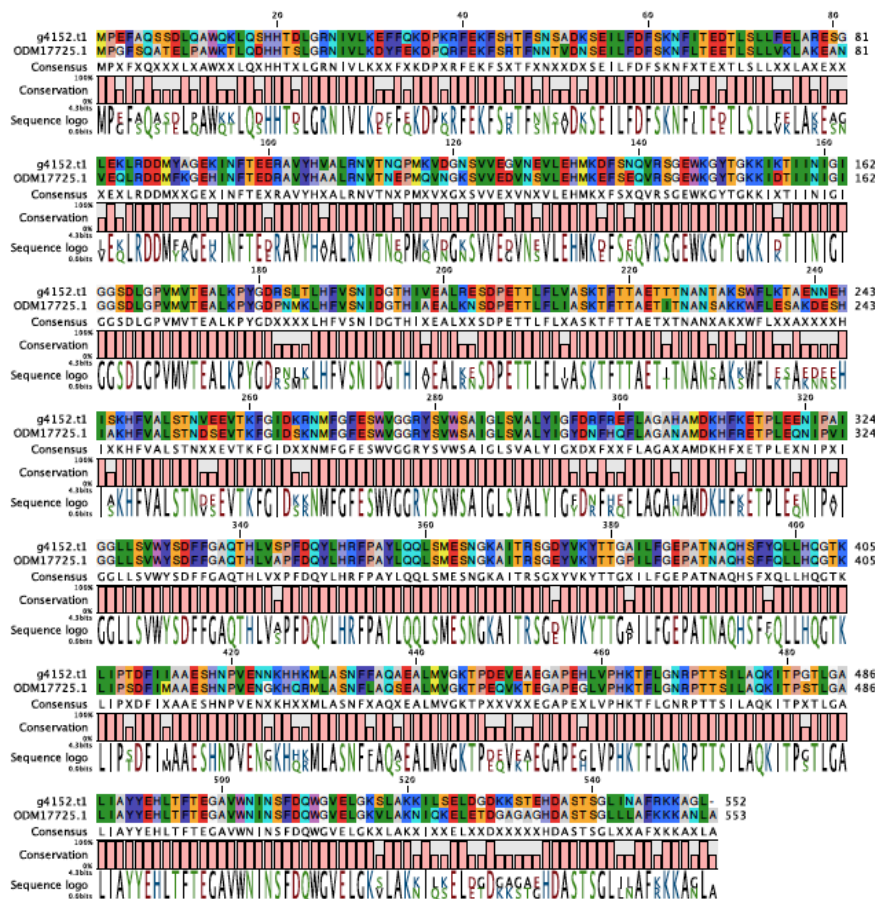


Figure S7: Sequence alignment of Glucose-6-phosphate isomerase

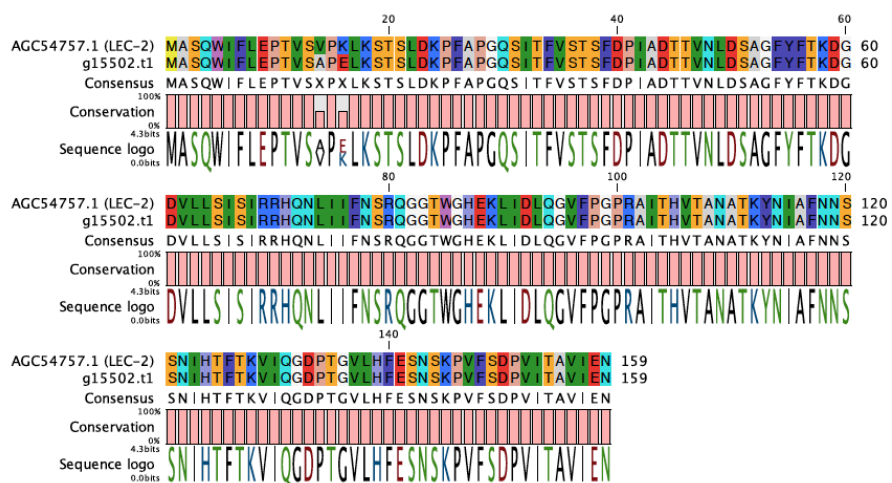
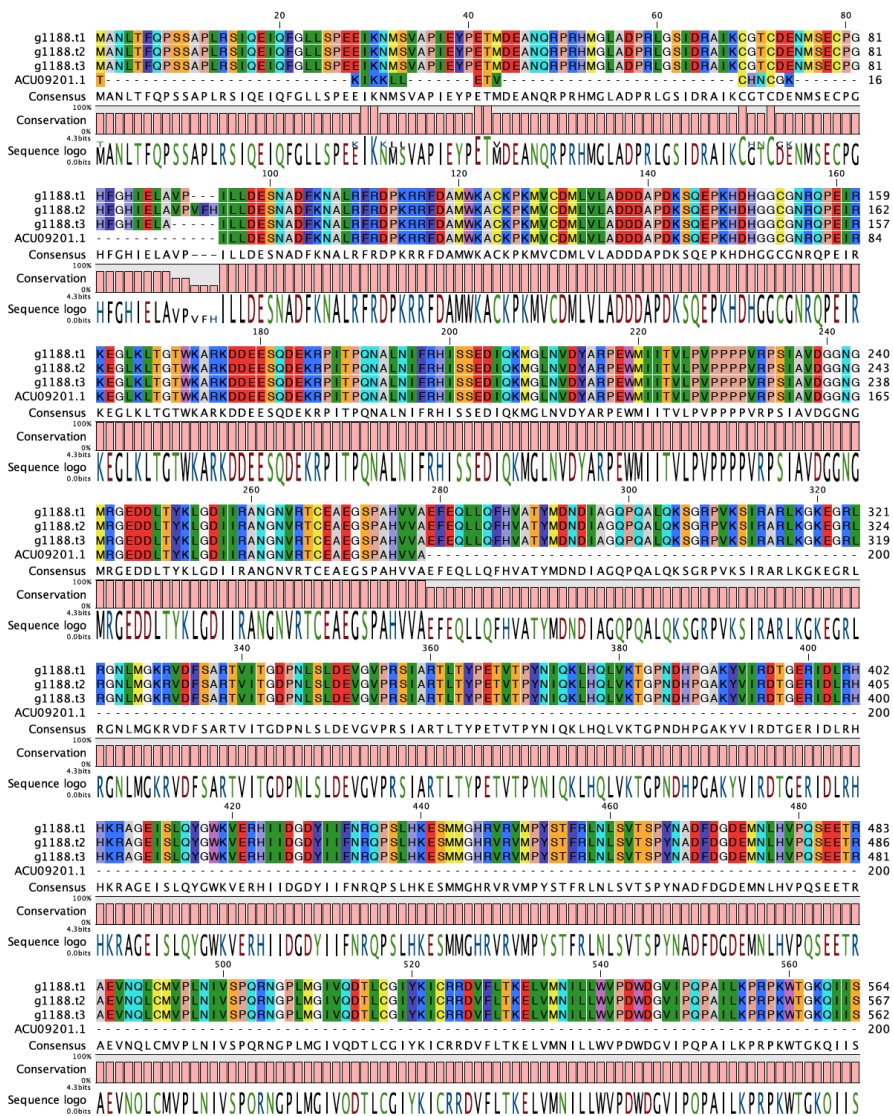


Figure S11: Sequence alignment of galectin like protein (*lec-2*)



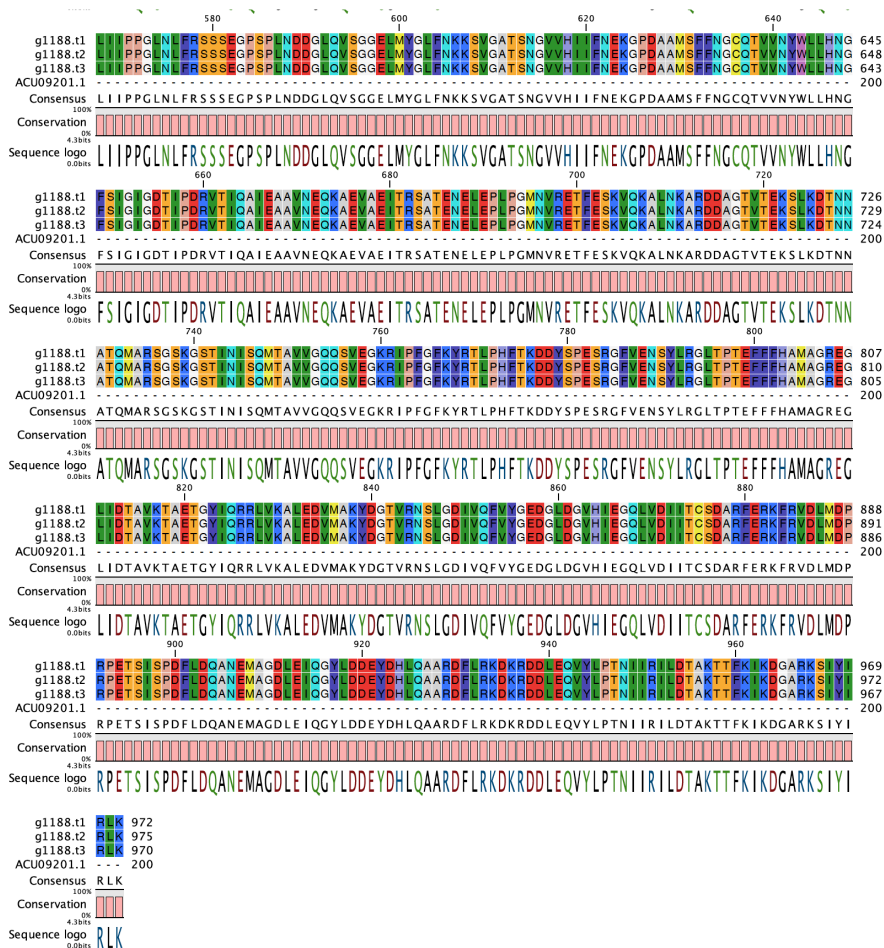


Figure S12: Sequence alignment of RNA polymerase II large subunit, partial.

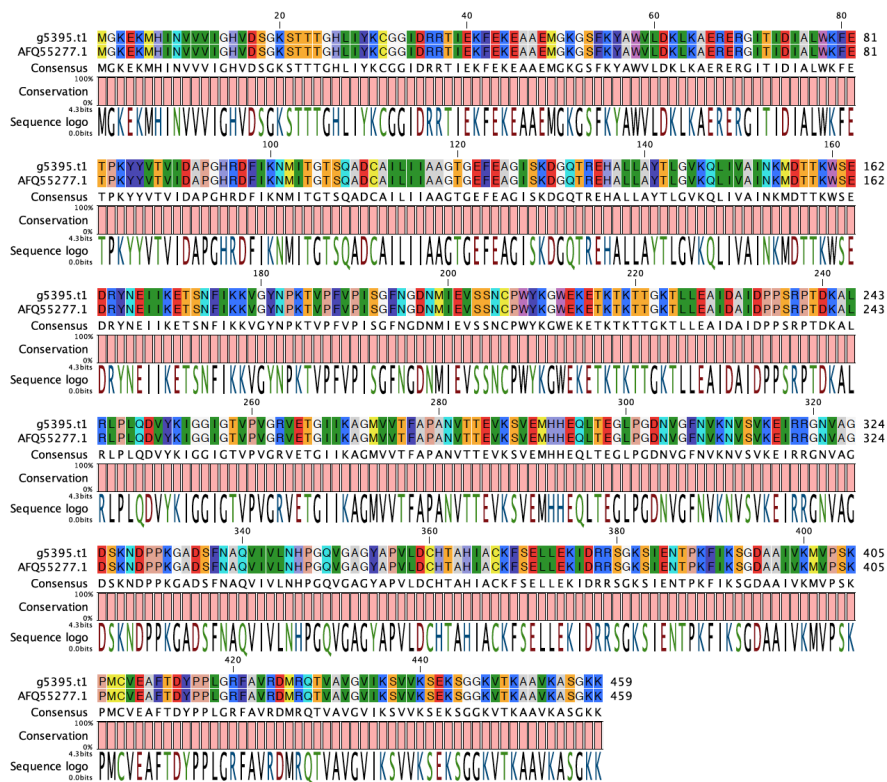


Figure S13: Sequence alignment of Elongation factor-1 alpha.



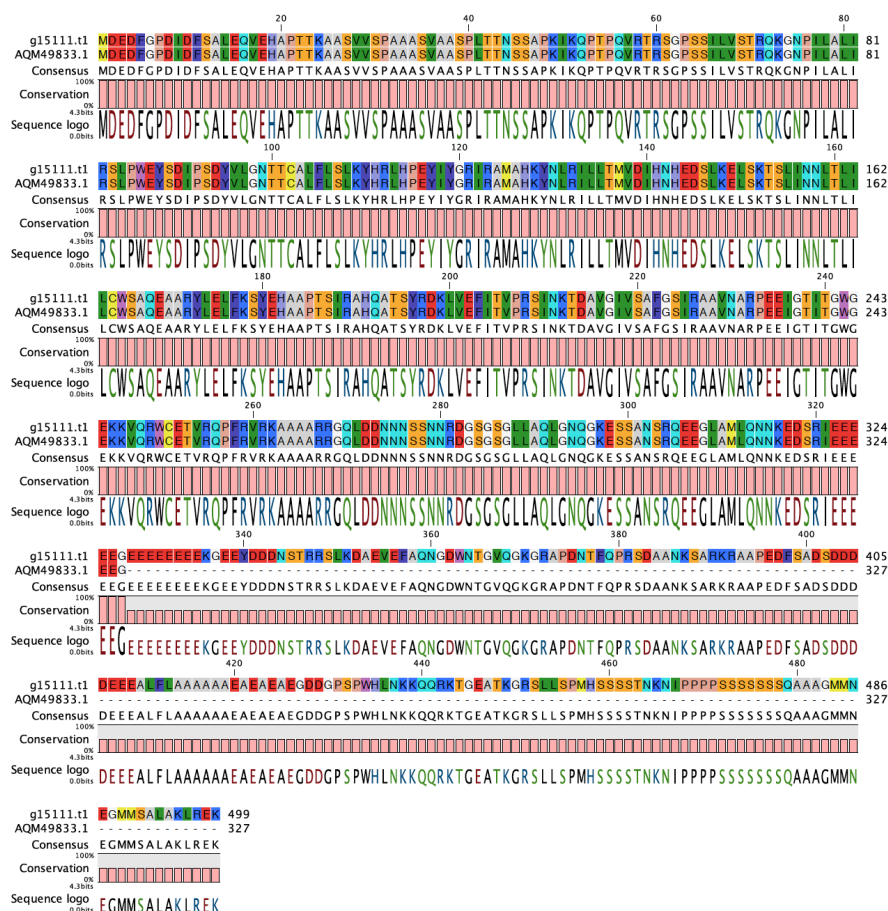


Figure S15: Sequence alignment of mating-type switching protein swi10, partial.



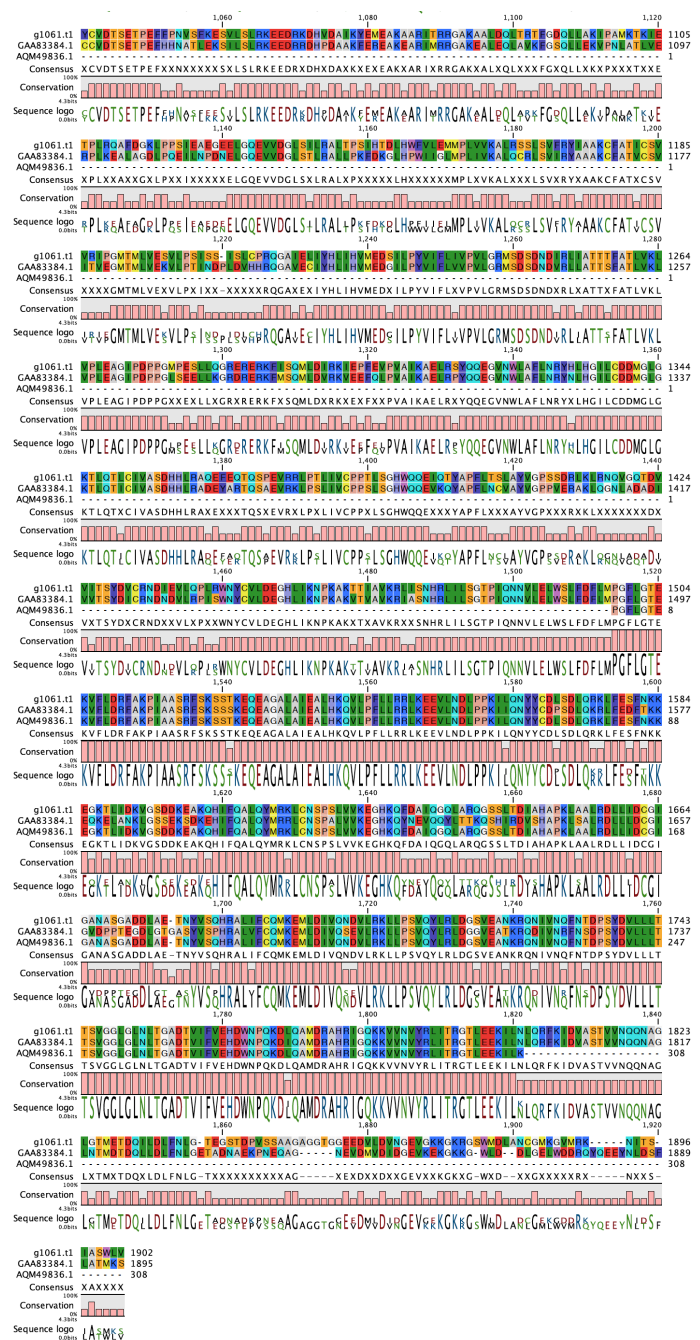


Figure S16: Sequence alignment of TBP (TATA binding protein) associated factor.

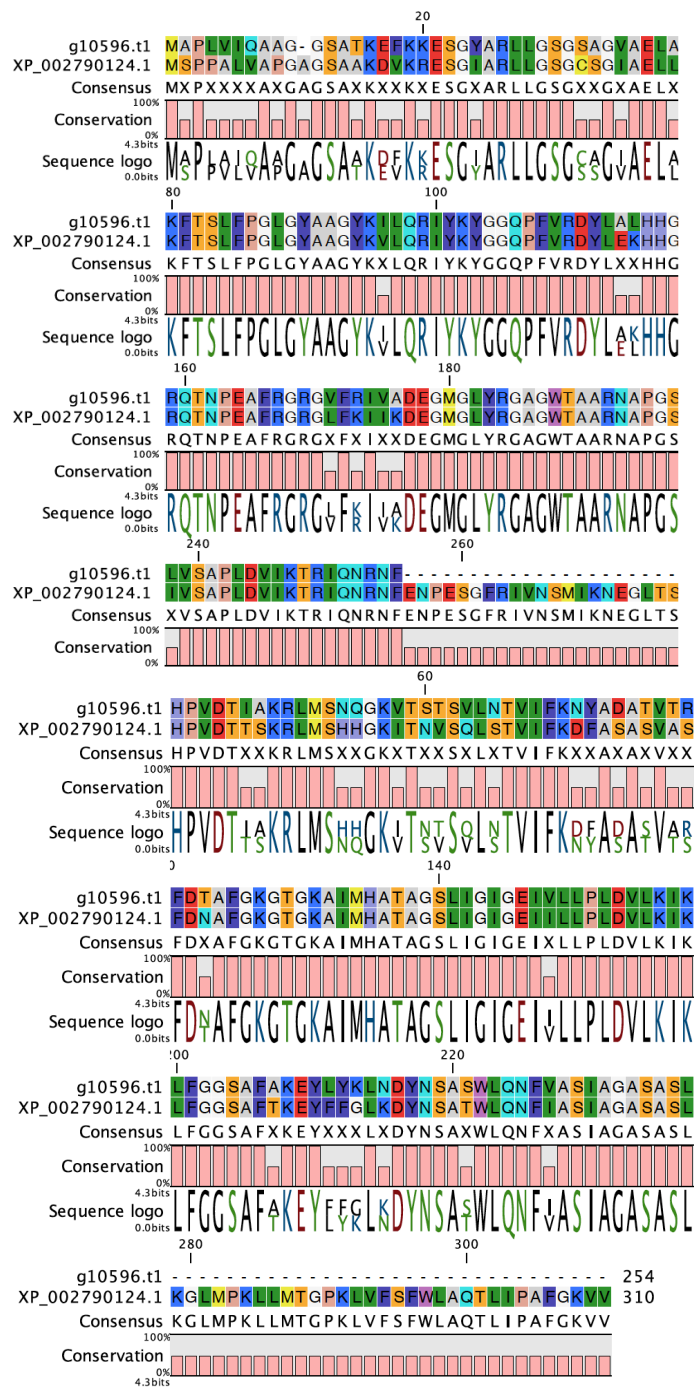


Figure S17: Sequence alignment of mitochondrial carrier protein.



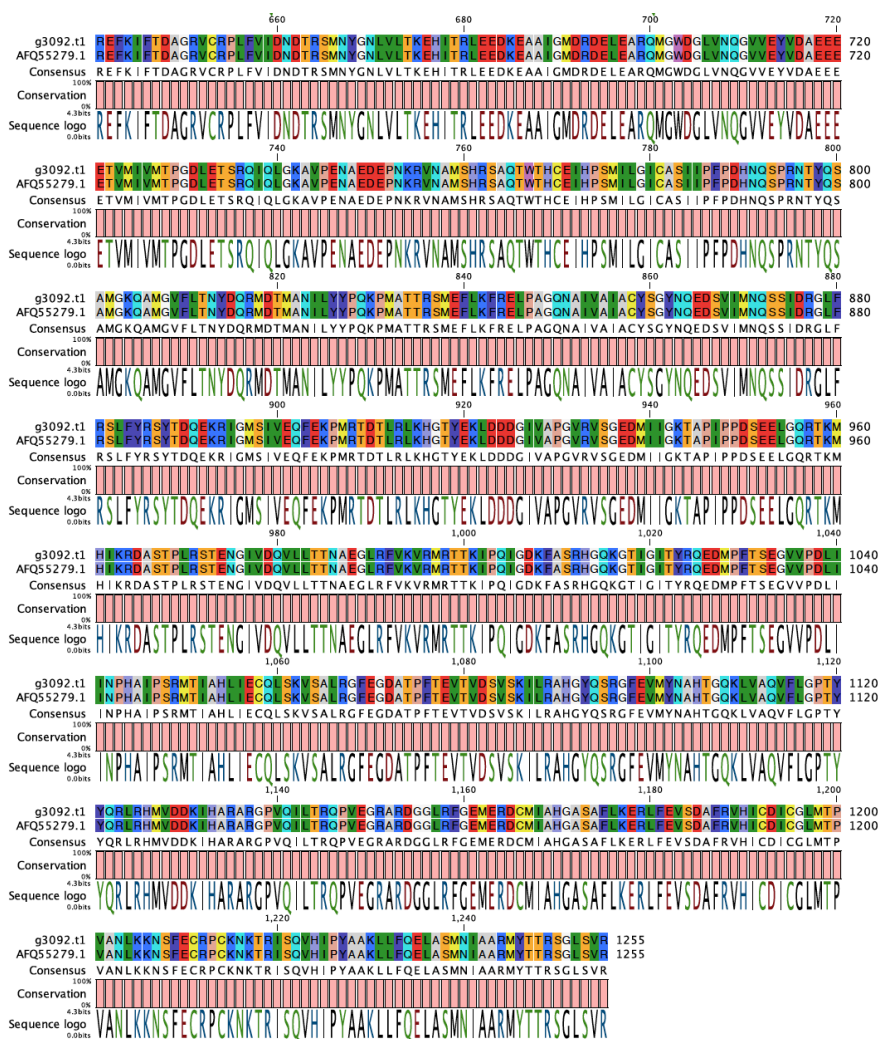


Figure S18: Sequence alignment of DNA directed RNA polymerase II.

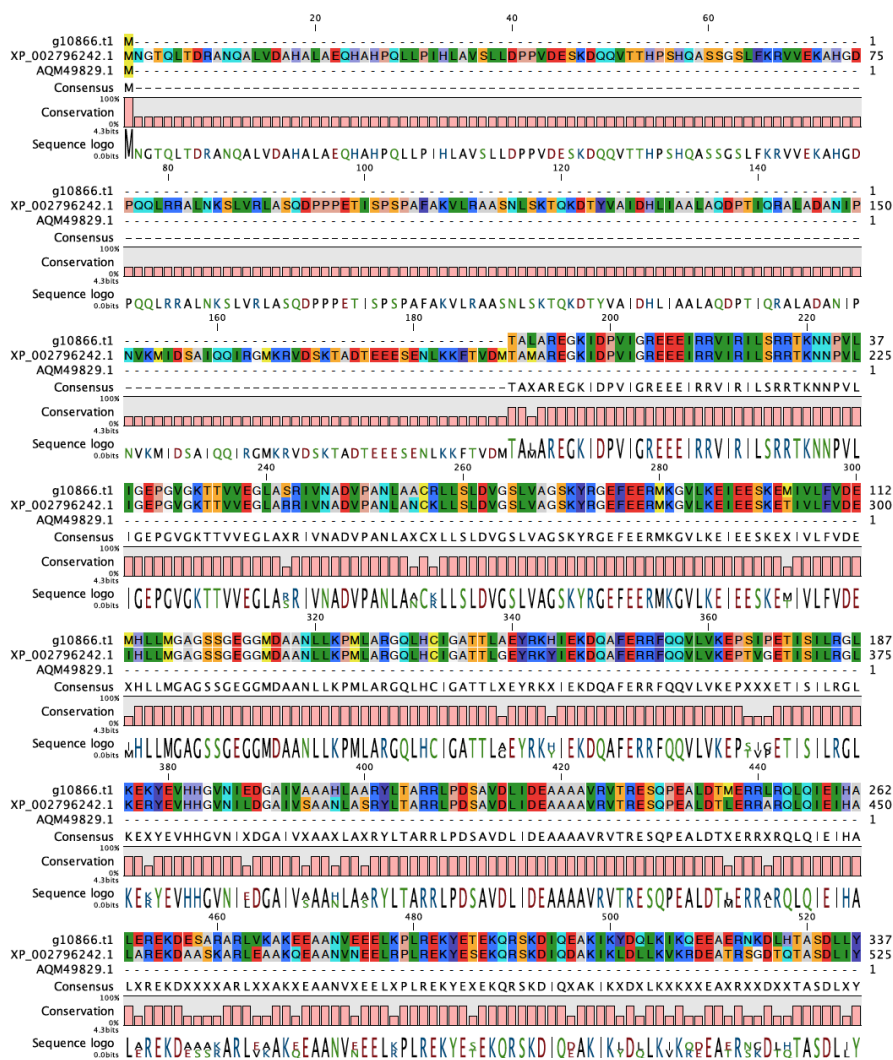


Figure S19: Sequence alignment of heat shock protein (*hsp98*).

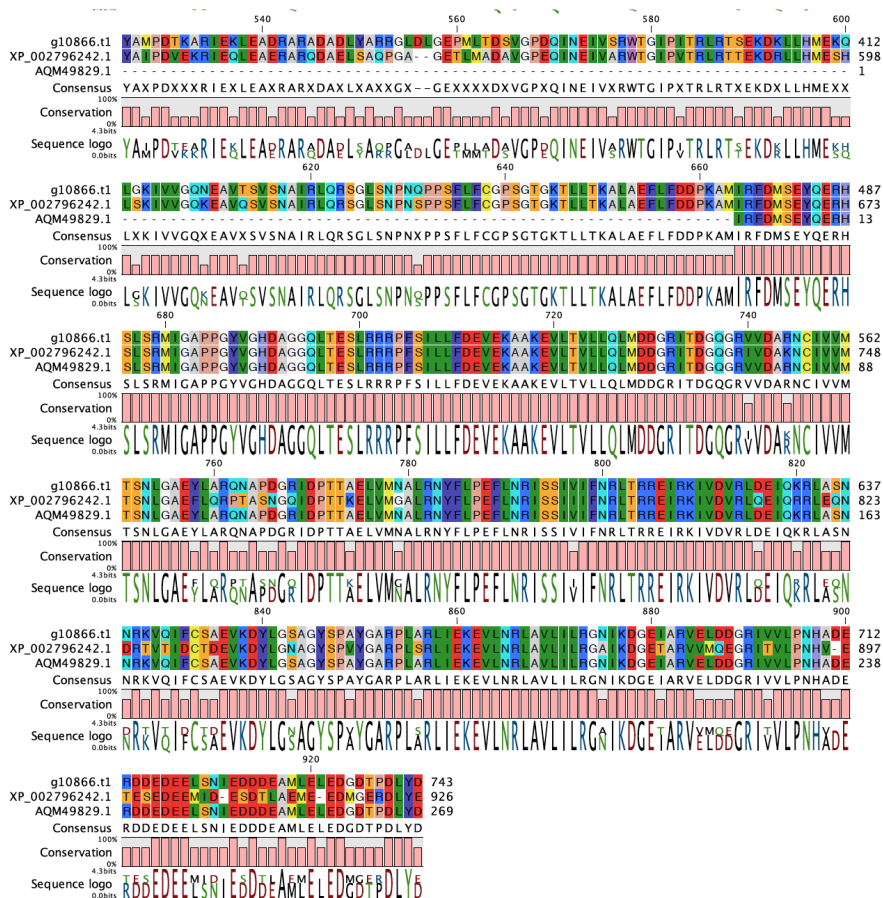


Figure S20: Sequence alignment of beta-tubulin (*tub-2*). Shaded amino acid sequences indicate similarities, and dashes indicate gaps introduced by the CLUSTALW (Thompson et al. 1994) to maximize alignment.

Appendix G: KEGG analysis.

Table S16: List of metabolic pathways removed from *P. membranacea*.

Pathway ID and Pathway	# Enzs identified
00195 Photosynthesis	7
00196 Photosynthesis - antenna proteins	5
00710 Carbon fixation in photosynthetic organisms	16
00720 Carbon fixation pathways in prokaryotes	17
00471 D-Glutamine and D-glutamate metabolism	1
00472 D-Arginine and D-ornithine metabolism	1
00532 Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	1
00981 Insect hormone biosynthesis	1
00908 Zeatin biosynthesis	1
01051 Biosynthesis of ansamycins	1
00941 Flavonoid biosynthesis	1
00944 Flavone and flavonol biosynthesis	1
00901 Indole alkaloid biosynthesis	1
00311 Penicillin and cephalosporin biosynthesis	1
00332 Carbapenem biosynthesis	2
00261 Monobactam biosynthesis	4
00521 Streptomycin biosynthesis	4
00524 Neomycin, kanamycin and gentamicin biosynthesis	1
00401 Novobiocin biosynthesis	3
00405 Phenazine biosynthesis	1
00333 Prodigiosin biosynthesis	2
00254 Aflatoxin biosynthesis	1
00364 Fluorobenzoate degradation	1
00621 Dioxin degradation	1
00624 Polycyclic aromatic hydrocarbon degradation	1

Table S17: The KEGG pathway table of *P. membranacea*.

Pathway module	Status
Energy metabolism	
Carbon fixation	
M00165 Reductive pentose phosphate cycle (Calvin cycle) [PATH:map01200 map00710] (12)	incomplete
M00166 Reductive pentose phosphate cycle, ribulose-5P => glyceraldehyde-3P [PATH:map01200 map00710] (3)	2 blocks missing
M00167 Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P [PATH:map01200 map00710] (9)	1 block missing
M00168 CAM (Crassulacean acid metabolism), dark [PATH:map01200 map00710] (2)	1 block missing
M00169 CAM (Crassulacean acid metabolism), light [PATH:map01200 map00710] (1)	1 block missing
M00172 C4-dicarboxylic acid cycle, NADP - malic enzyme type [PATH:map01200 map00710] (1)	incomplete
M00171 C4-dicarboxylic acid cycle, NAD - malic enzyme type [PATH:map01200 map00710] (9)	incomplete
M00170 C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type [PATH:map01200 map00710] (8)	1 block missing
M00173 Reductive citrate cycle (Arnon-Buchanan cycle) [PATH:map01200 map00720] (4)	incomplete
M00376 3-Hydroxypropionate bi-cycle [PATH:map01200 map00720] (1)	incomplete
M00375 Hydroxypropionate-hydroxybutylate cycle [PATH:map01200 map00720] (2)	incomplete
M00374 Dicarboxylate-hydroxybutyrate cycle [PATH:map01200 map00720] (2)	incomplete
M00377 Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) [PATH:map01200 map00720] (3)	incomplete
Nitrogen metabolism	
M00531 Assimilatory nitrate reduction, nitrate => ammonia [PATH:map00910] (2)	complete

Methane metabolism		
M00357 Methanogenesis, acetate => methane [PATH:map01200 map00680] (1)		incomplete
M00346 Formaldehyde assimilation, serine pathway [PATH:map01200 map00680] (3)		incomplete
M00345 Formaldehyde assimilation, ribulose monophosphate pathway [PATH:map01200 map00680] (4)		1 block missing
M00344 Formaldehyde assimilation, xylulose monophosphate pathway [PATH:map01200 map00680] (3)		2 blocks missing
Sulfur metabolism		
M00176 Assimilatory sulfate reduction, sulfate => H ₂ S [PATH:map00920] (8)		complete
M00596 Dissimilatory sulfate reduction, sulfate => H ₂ S [PATH:map00920] (1)		2 blocks missing
M00595 Thiosulfate oxidation by SOX complex, thiosulfate => sulfate [PATH:map00920] (1)		1 block missing
Carbohydrate and lipid metabolism		
Central carbohydrate metabolism		
M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate [PATH:map01200 map00010] (19)		1 block missing
M00002 Glycolysis, core module involving three-carbon compounds [PATH:map01200 map01230 map00010] (10)		1 block missing
M00003 Gluconeogenesis, oxaloacetate => fructose-6P [PATH:map00010 map00020] (14)		1 block missing
M00307 Pyruvate oxidation, pyruvate => acetyl-CoA [PATH:map01200 map00010 map00020 map00620] (5)		complete
M00009 Citrate cycle (TCA cycle, Krebs cycle) [PATH:map01200 map00020] (18)		complete
M00010 Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate [PATH:map01200 map01210 map01230 map00020] (6)		complete
M00011 Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate [PATH:map01200 map00020] (12)		complete
M00004 Pentose phosphate pathway (Pentose phosphate cycle)		complete

[PATH:map01200 map00030] (15)	
M00006 Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	
[PATH:map01200 map00030] (4)	complete
M00007 Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	
[PATH:map01200 map01230 map00030] (10)	complete
M00580 Pentose phosphate pathway, archaea, fructose 6P => ribose 5P	
[PATH:map01200 map01230 map00030] (1)	1 block missing
M00005 PRPP biosynthesis, ribose 5P => PRPP	
[PATH:map01200 map01230 map00030 map00230] (3)	complete
M00008 Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	
[PATH:map01200 map00030] (2)	2 blocks missing
M00308 Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glycerate-3P	
[PATH:map01200 map00030] (3)	incomplete
Other carbohydrate metabolism	
M00012 Glyoxylate cycle [PATH:map01200 map00630] (8)	complete
M00373 Ethylmalonyl pathway [PATH:map01200 map00630] (2)	incomplete
M00740 Methylaspartate cycle [PATH:map01200 map00020 map00250 map00630] (4)	incomplete
M00532 Photorespiration [PATH:map01200 map00630] (10)	incomplete
M00013 Malonate semialdehyde pathway, propanoyl-CoA => acetyl-CoA	
[PATH:map01200 map00640] (5)	complete
M00632 Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	
[PATH:map00052] (5)	complete
M00552 D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	
[PATH:map00052] (3)	incomplete
M00014 Glucuronate pathway (uronate pathway) [PATH:map00040] (3)	incomplete
M00565 Trehalose biosynthesis, D-glucose 1P => trehalose [PATH:map00500] (1)	incomplete

M00549 Nucleotide sugar biosynthesis, glucose => UDP-glucose [PATH:map00520] (7)	complete
M00554 Nucleotide sugar biosynthesis, galactose => UDP-galactose [PATH:map00520 map00052] (2)	complete
Fatty acid metabolism	
M00082 Fatty acid biosynthesis, initiation [PATH:map01212 map00061] (5)	complete
M00083 Fatty acid biosynthesis, elongation [PATH:map01212 map00061] (6)	complete
M00085 Fatty acid biosynthesis, elongation, mitochondria [PATH:map01212 map00062] (2)	2 Blocks missing
M00415 Fatty acid biosynthesis, elongation, endoplasmic reticulum [PATH:map01212 map00062] (4)	complete
M00113 Jasmonic acid biosynthesis [PATH:map00592] (2)	incomplete
M00086 beta-Oxidation, acyl-CoA synthesis [PATH:map01212 map00071] (3)	complete
M00087 beta-Oxidation [PATH:map01212 map00071] (4)	1 block missing
M00088 Ketone body biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone [PATH:map00072] (4)	2 blocks missing
Lipid metabolism	
M00089 Triacylglycerol biosynthesis [PATH:map00561] (5)	1 block missing
M00098 Acylglycerol degradation [PATH:map00561] (7)	complete
M00090 Phosphatidylcholine (PC) biosynthesis, choline => PC [PATH:map00564] (3)	1 block missing
M00091 Phosphatidylcholine (PC) biosynthesis, PE => PC [PATH:map00564] (1)	1 block missing
M00092 Phosphatidylethanolamine (PE) biosynthesis, ethanolamine => PE [PATH:map00564] (5)	complete
M00093 Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE [PATH:map00564] (4)	complete
M00130 Inositol phosphate metabolism, PI=> PIP2 => Ins(1,4,5)P3 => Ins(1,3,4,5)P4 [PATH:map00562] (5)	1 block missing
M00131 Inositol phosphate metabolism, Ins(1,3,4,5)P4 => Ins(1,3,4)P3 => myo-inositol [PATH:map00562] (1)	incomplete
M00132 Inositol phosphate metabolism, Ins(1,3,4)P3 => phytate [PATH:map00562] (1)	2 blocks missing

M00094 Ceramide biosynthesis [PATH:map00600] (7)	complete
M00066 Lactosylceramide biosynthesis [PATH:map00600] (1)	1 block missing
M00099 Sphingosine biosynthesis [PATH:map00600] (8)	complete
M00100 Sphingosine degradation [PATH:map00600] (2)	complete
Glycan metabolism	
M00055 N-glycan precursor biosynthesis [PATH:map00510] (14)	1 block missing
M00072 N-glycosylation by oligosaccharyltransferase [PATH:map00510] (6)	complete
M00073 N-glycan precursor trimming [PATH:map00510] (8)	complete
M00074 N-glycan biosynthesis, high-mannose type [PATH:map00510 map00513] (14)	2 blocks missing
M00075 N-glycan biosynthesis, complex type [PATH:map00510] (2)	incomplete
M00065 GPI-anchor biosynthesis, core oligosaccharide [PATH:map00563] (6)	2 blocks missing
Glycosaminoglycan metabolism	
M00079 Keratan sulfate degradation [PATH:map00531] (2)	incomplete
Terpenoid backbone biosynthesis	
M00095 C5 isoprenoid biosynthesis, mevalonate pathway [PATH:map00900] (8)	complete
M00096 C5 isoprenoid biosynthesis, non-mevalonate pathway [PATH:map00900] (1)	incomplete
M00364 C10-C20 isoprenoid biosynthesis, bacteria [PATH:map00900] (1)	1 block missing
M00365 C10-C20 isoprenoid biosynthesis, archaea [PATH:map00900] (1)	1 block missing
M00366 C10-C20 isoprenoid biosynthesis, plants [PATH:map00900] (2)	2 blocks missing
M00367 C10-C20 isoprenoid biosynthesis, non-plant eukaryotes [PATH:map00900] (4)	complete
Sterol biosynthesis	
M00101 Cholesterol biosynthesis, squalene 2,3-epoxide => cholesterol [PATH:map00100] (11)	incomplete
M00102 Ergocalciferol biosynthesis [PATH:map00100] (4)	1 block missing
M00104 Bile acid biosynthesis, cholesterol => cholate [PATH:map00120] (2)	incomplete
Nucleotide and amino acid metabolism	
Purine metabolism	

M00048 Inosine monophosphate biosynthesis, PRPP + glutamine => IMP [PATH:map00230] (12)	complete
M00049 Adenine ribonucleotide biosynthesis, IMP => ADP,ATP [PATH:map00230] (8)	complete
M00050 Guanine ribonucleotide biosynthesis IMP => GDP,GTP [PATH:map00230] (5)	complete
M00546 Purine degradation, xanthine => urea [PATH:map00230] (6)	1 block missing
Pyrimidine metabolism	
M00051 Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP [PATH:map00240] (10)	complete
M00052 Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP [PATH:map00240] (4)	complete
M00053 Pyrimidine deoxyribonucleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP [PATH:map00240] (7)	2 blocks missing
Serine and threonine metabolism	
M00020 Serine biosynthesis, glycerate-3P => serine [PATH:map01200 map01230 map00260] (5)	complete
M00018 Threonine biosynthesis, aspartate => homoserine => threonine [PATH:map01230 map00260] (5)	complete
Cysteine and methionine metabolism	
M00021 Cysteine biosynthesis, serine => cysteine [PATH:map01200 map01230 map00270] (3)	1 block missing
M00338 Cysteine biosynthesis, homocysteine + serine => cysteine [PATH:map01230 map00270] (2)	complete
M00609 Cysteine biosynthesis, methionine => cysteine [PATH:map01230 map00270] (4)	incomplete
M00017 Methionine biosynthesis, aspartate => homoserine => methionine [PATH:map01230 map00270] (6)	1 block missing
M00034 Methionine salvage pathway [PATH:map00270] (12)	1 block missing
M00035 Methionine degradation [PATH:map00270] (8)	complete
M00368 Ethylene biosynthesis, methionine => ethylene [PATH:map00270] (4)	2 blocks missing
Branched-chain amino acid metabolism	

M00019 Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine [PATH:map01210 map01230 map00290] (12)	complete
M00535 Isoleucine biosynthesis, pyruvate => 2-oxobutanoate [PATH:map01210 map01230 map00290] (2)	2 blocks missing
M00570 Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine [PATH:map01230 map00290] (15)	complete
M00432 Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate [PATH:map01210 map01230 map00290] (7)	complete
M00036 Leucine degradation, leucine => acetoacetate + acetyl-CoA [PATH:map00280] (16)	complete

Lysine metabolism

M00016 Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine [PATH:map01230 map00300] (2)	incomplete
M00525 Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine [PATH:map01230 map00300] (2)	incomplete
M00526 Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine [PATH:map01230 map00300] (2)	incomplete
M00527 Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine [PATH:map01230 map00300] (2)	incomplete
M00030 Lysine biosynthesis, AAA pathway, 2-oxoglutarate => 2-aminoadipate => lysine [PATH:map01230 map00300] (13)	complete
M00433 Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate [PATH:map01210 map01230 map00300] (7)	complete
M00032 Lysine degradation, lysine => saccharopine => acetoacetyl-CoA [PATH:map00310] (7)	incomplete

Arginine and proline metabolism

M00015 Proline biosynthesis, glutamate => proline [PATH:map01230 map00330] (3)	complete
M00028 Ornithine biosynthesis, glutamate => ornithine	complete

[PATH:map01210 map01230 map00220] (8)	
M00029 Urea cycle [PATH:map01230 map00220] (5)	2 blocks missing
Histidine metabolism	
M00026 Histidine biosynthesis, PRPP => histidine [PATH:map01230 map00340] (14)	complete
Aromatic amino acid metabolism	
M00022 Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate [PATH:map01230 map00400] (7)	complete
M00023 Tryptophan biosynthesis, chorismate => tryptophan [PATH:map01230 map00400] (8)	complete
M00024 Phenylalanine biosynthesis, chorismate => phenylalanine [PATH:map01230 map00400] (3)	complete
M00025 Tyrosine biosynthesis, chorismate => tyrosine [PATH:map01230 map00400] (2)	1 block missing
M00042 Catecholamine biosynthesis, tyrosine => dopamine => noradrenaline => adrenaline [PATH:map00350] (15)	2 blocks missing
M00044 Tyrosine degradation, tyrosine => homogentisate [PATH:map00350] (2)	incomplete
M00350 Capsaicin biosynthesis, L-Phenylalanine => Capsaicin [PATH:map00360] (2)	incomplete
M00037 Melatonin biosynthesis, tryptophan => serotonin => melatonin [PATH:map00380] (1)	incomplete
M00038 Tryptophan metabolism, tryptophan => kynurenine => 2-aminomuconate [PATH:map00380] (4)	incomplete
Other amino acid metabolism	
M00033 Ectoine biosynthesis, aspartate => ectoine [PATH:map01210 map01230 map00260] (2)	incomplete
M00027 GABA (gamma-Aminobutyrate) shunt [PATH:map00250] (4)	complete
M00047 Creatine pathway [PATH:map00330] (2)	2 blocks missing
Cofactor and vitamin biosynthesis	
M00127 Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P [PATH:map00730] (1)	incomplete
M00125 Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD [PATH:map00740] (6)	2 blocks missing
M00124 Pyridoxal biosynthesis, erythrose-4P => pyridoxal-5P [PATH:map00750] (3)	incomplete

M00115 NAD biosynthesis, aspartate => NAD [PATH:map00760] (3)	2 blocks missing
M00119 Pantothenate biosynthesis, valine/L-aspartate => pantothenate [PATH:map00770] (8)	1 block missing
M00120 Coenzyme A biosynthesis, pantothenate => CoA [PATH:map00770] (4)	1 block missing
M00572 Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP [PATH:map00780] (4)	incomplete
M00126 Tetrahydrofolate biosynthesis, GTP => THF [PATH:map00790 map00670] (7)	1 block missing
M00121 Heme biosynthesis, glutamate => protoheme/siroheme [PATH:map00860] (13)	incomplete
M00129 Ascorbate biosynthesis, animals, glucose-1P => ascorbate [PATH:map00040 map00053] (2)	incomplete
M00114 Ascorbate biosynthesis, plants, glucose-6P => ascorbate [PATH:map00010 map00051 map00053] (6)	incomplete
M00117 Ubiquinone biosynthesis, prokaryotes, chorismate => ubiquinone [PATH:map00130] (2)	incomplete
M00128 Ubiquinone biosynthesis, eukaryotes, 4-hydroxybenzoate => ubiquinone [PATH:map00130] (4)	2 blocks missing
M00118 Glutathione biosynthesis, glutamate => glutathione [PATH:map00480] (5)	complete
M00140 C1-unit interconversion, prokaryotes [PATH:map00670] (2)	2 blocks missing
M00141 C1-unit interconversion, eukaryotes [PATH:map00670] (3)	complete
Polyamine biosynthesis	
M00133 Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine [PATH:map00330] (3)	1 block missing
M00134 Polyamine biosynthesis, arginine => ornithine => putrescine [PATH:map00330] (5)	complete
M00135 GABA biosynthesis, eukaryotes, putrescine => GABA [PATH:map00330] (1)	2 blocks missing
Secondary metabolism	
Biosynthesis of secondary metabolites	
M00039 Monolignol biosynthesis, phenylalanine / tyrosine => monolignol [PATH:map00940] (2)	incomplete
M00137 Flavanone biosynthesis, phenylalanine => naringenin [PATH:map00940 map00941] (2)	incomplete

Structural complex

Energy metabolism

Photosynthesis

M00161 Photosystem II [PATH:map00195] [BR:ko00194] (1)	1 block missing
M00163 Photosystem I [PATH:map00195] [BR:ko00194] (2)	1 block missing

ATP synthesis

M00145 NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria [PATH:map00190] (1)	1 block missing
M00143 NADH dehydrogenase (ubiquinone) Fe-S protein/flavoprotein complex, mitochondria [PATH:map00190] (10)	complete
M00146 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex [PATH:map00190] (5)	1 block missing
M00147 NADH dehydrogenase (ubiquinone) 1 beta subcomplex [PATH:map00190] (3)	1 block missing
M00148 Succinate dehydrogenase (ubiquinone) [PATH:map00190] (4)	complete
M00151 Cytochrome bc1 complex respiratory unit [PATH:map00190] (2)	1 block missing
M00152 Cytochrome bc1 complex [PATH:map00190] (6)	1 block missing
M00154 Cytochrome c oxidase [PATH:map00190] (10)	1 block missing
M00155 Cytochrome c oxidase, prokaryotes [PATH:map00190] (1)	1 block missing
M00157 F-type ATPase, prokaryotes and chloroplasts [PATH:map00190 map00195] (3)	1 block missing
M00158 F-type ATPase, eukaryotes [PATH:map00190] (14)	1 block missing
M00160 V-type ATPase, eukaryotes [PATH:map00190] (14)	1 block missing

Genetic information processing

DNA polymerase

M00261 DNA polymerase alpha / primase complex [PATH:map03030] [BR:ko03032] (4)	complete
M00262 DNA polymerase delta complex [PATH:map03030 map03410 map03420]	complete

[BR:ko03032 ko03400] (5)	
M00263 DNA polymerase epsilon complex [PATH:map03030] [BR:ko03032 ko03400] (3)	1 block missing
M00293 DNA polymerase zeta complex [PATH:map03460] [BR:ko03400] (3)	complete
M00294 DNA polymerase gamma complex [BR:ko03032] (1)	1 block missing
Replication system	
M00284 Origin recognition complex [PATH:map04110 map04111 map04113] [BR:ko03032] (6)	complete
M00285 MCM complex [PATH:map03030 map04110 map04111 map04113] [BR:ko03032] (7)	complete
M00286 GINS complex [BR:ko03032] (3)	1 block missing
M00288 RPA complex [PATH:map03030 map03420 map03430 map03440] [BR:ko03032 ko03400] (3)	complete
M00289 RF-C complex [PATH:map03030 map03420 map03430] [BR:ko03032 ko03400] (6)	complete
Repair system	
M00296 BER complex [PATH:map03410] [BR:ko03400] (5)	1 block missing
M00290 Holo-TFIIH complex [PATH:map03420] [BR:ko03021 ko03400] (10)	complete
M00291 MRN complex [PATH:map03440] [BR:ko03032 ko03400] (2)	1 block missing
M00292 MRX complex [PATH:map03440 map03450] [BR:ko03400] (2)	1 block missing
M00297 DNA-PK complex [PATH:map03450 map04110] [BR:ko03032 ko03400] (2)	1 block missing
M00295 BRCA1-associated genome surveillance complex (BASC) [PATH:map03430 map03440 map03460] [BR:ko03400] (18)	1 block missing
M00414 Bloom's syndrome complex [PATH:map03460] [BR:ko03400] (4)	1 block missing
RNA polymerase	
M00183 RNA polymerase, bacteria [PATH: map03020] [BR: ko03021 ko03400] (1)	2 blocks missing
M00180 RNA polymerase II, eukaryotes [PATH:map03020 map05016] [BR:ko03021 ko03400] (13)	complete
M00181 RNA polymerase III, eukaryotes [PATH:map03020 map04623] [BR:ko03021] (17)	complete
M00182 RNA polymerase I, eukaryotes [PATH:map03020] [BR:ko03021] (12)	complete

Spliceosome	
M00351 Spliceosome, U1-snRNP [PATH:map03040] [BR:ko03041] (8)	1 block missing
M00352 Spliceosome, U2-snRNP [PATH:map03040] [BR:ko03041] (19)	1 block missing
M00354 Spliceosome, U4/U6.U5 tri-snRNP [PATH:map03040] [BR:ko03041] (35)	complete
M00355 Spliceosome, 35S U5-snRNP [PATH:map03040] [BR:ko03041] (34)	complete
M00353 Spliceosome, Prp19/CDC5L complex [PATH:map03040] [BR:ko03041] (9)	1 block missing
M00396 Lsm 2-8 complex [PATH:map03040 map03018] [BR:ko03041] (7)	complete
M00397 Lsm 1-7 complex [PATH:map03018] (7)	complete
M00398 Sm core complex [PATH:map03040] [BR:ko03032 ko03041] (6)	complete
M00399 Cap binding complex [PATH:map03013 map03015 map03040 map03040] [BR:ko03041] (2)	complete
RNA processing	
M00425 H/ACA ribonucleoprotein complex [PATH:map03008] [BR:ko03032 ko03009] (4)	complete
M00426 Survival motor neuron (SMN) complex [PATH:map03013] [BR:ko03041] (1)	1 block missing
M00427 Nuclear pore complex [PATH:map03013] (25)	1 block missing
M00428 eIF4F complex [PATH:map03013] [BR:ko03012] (4)	complete
M00430 Exon junction complex (EJC) [PATH:map03013 map03015 map03040] [BR:ko03041] (17)	1 block missing
M00405 THC complex [PATH:map03040 map03013] [BR:ko03041] (3)	1 block missing
M00406 TREX complex [PATH:map03013 map03040] [BR:ko03041] (7)	1 block missing
M00390 Exosome, archaea [PATH:map03018] (4)	complete
M00391 Exosome, eukaryotes [PATH:map03018] (10)	complete
M00392 Ski complex [PATH:map03018] (6)	complete
M00393 TRAMP complex [PATH:map03018] (3)	1 block missing
M00395 Decapping complex [PATH:map03018] (6)	1 block missing
Ribosome	

M00178 Ribosome, bacteria [PATH:map03010] [BR:ko03011] (25)	1 block missing
M00179 Ribosome, archaea [PATH:map03010] [BR:ko03011] (49)	1 block missing
M00177 Ribosome, eukaryotes [PATH:map03010] [BR:ko03011] (72)	1 block missing
Proteasome	
M00340 Proteasome, 20S core particle [PATH:map03050] [BR:ko03051] (14)	complete
M00341 Proteasome, 19S regulatory particle (PA700) [PATH:map03050] [BR:ko03051] (19)	complete
M00337 Immunoproteasome [PATH:map03050] [BR:ko03051] (11)	1 block missing
M00342 Bacterial proteasome [PATH:map03050] [BR:ko03051] (1)	1 block missing
Ubiquitin system	
M00379 SCF-MET30 complex [PATH:map04120 map04111] [BR:ko04121] (5)	complete
M00380 SCF-BTRC complex [PATH:map04120 map04114 map04310 map04710] [BR:ko04121] (7)	complete
M00381 SCF-SKP2 complex [PATH:map04120 map04110 map05168] [BR:ko04121] (4)	1 block missing
M00382 SCF-FBS complex [PATH:map04120 map04141] [BR:ko04121] (4)	1 block missing
M00407 SCF-CDC4 complex [PATH:map04120 map04111] [BR:ko04121] (7)	complete
M00411 SCF-GRR1 complex [PATH:map04120 map04111] [BR:ko04121] (5)	complete
M00383 ECV complex [PATH:map04120 map05211] [BR:ko04121] (2)	1 block missing
M00384 Cul3-SPOP complex [PATH:map04120] [BR:ko04121] (2)	1 block missing
M00385 Cul4-DDB1-DDB2 complex [PATH:map04120 map03420] [BR:ko04121 ko03400] (4)	1 block missing
M00386 Cul4-DDB1-CSA complex [PATH:map04120 map03420] [BR:ko04121 ko03400] (5)	complete
M00387 SCF-FBW7 complex [PATH:map04120] [BR:ko04121] (4)	1 block missing
M00388 ECS complex [PATH:map04120] [BR:ko04121] (1)	1 block missing
M00389 APC/C complex [PATH:map04120 map04110 map04114 map04914 map04111 map04113] [BR:ko04121 ko03036] (13)	1 block missing
Protein processing	

M00400 p97-Ufd1-Npl4 complex [PATH:map04141] (4)	complete
M00401 Sec61 complex [PATH:map03060 map04141 map04145 map05110] [BR:ko02044] (4)	complete
M00403 HRD1/SEL1 ERAD complex [PATH:map04141] (10)	1 block missing
M00404 COPII complex [PATH:map04141] (8)	complete
M00408 ESCRT-0 complex [PATH:map04144] (3)	complete
M00409 ESCRT-I complex [PATH:map04144] (2)	1 block missing
M00410 ESCRT-II complex [PATH:map04144] (2)	1 block missing
M00412 ESCRT-III complex [PATH:map04144] (9)	complete

Environmental information processing

Metallic cation, iron-siderophore and vitamin B12 transport system

M00240 Iron complex transport system [PATH:map02010] [BR:ko02000] (1)	1 block missing
---	-----------------

Bacterial secretion system

M00335 Sec (secretion) system [PATH:map03070] [BR:ko02044] (2)	1 block missing
M00429 Competence-related DNA transformation transporter [BR:ko02044] (1)	1 block missing

Functional set

Metabolism

Aminoacyl tRNA

M00360 Aminoacyl-tRNA biosynthesis, prokaryotes [PATH: map00970] (53)	complete
M00359 Aminoacyl-tRNA biosynthesis, eukaryotes [PATH:map00970] (53)	complete

Nucleotide sugar

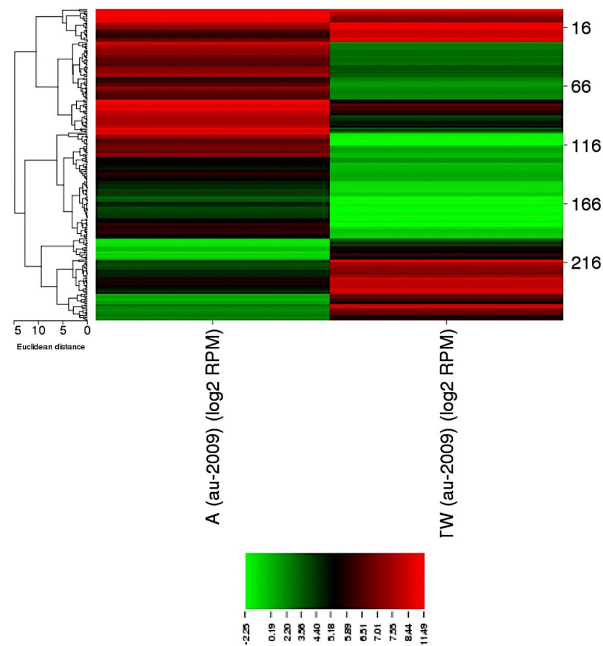
M00362 Nucleotide sugar biosynthesis, prokaryotes [PATH:map00520] (8)	1 block missing
M00361 Nucleotide sugar biosynthesis, eukaryotes [PATH:map00520] (7)	1 block missing

Environmental information processing

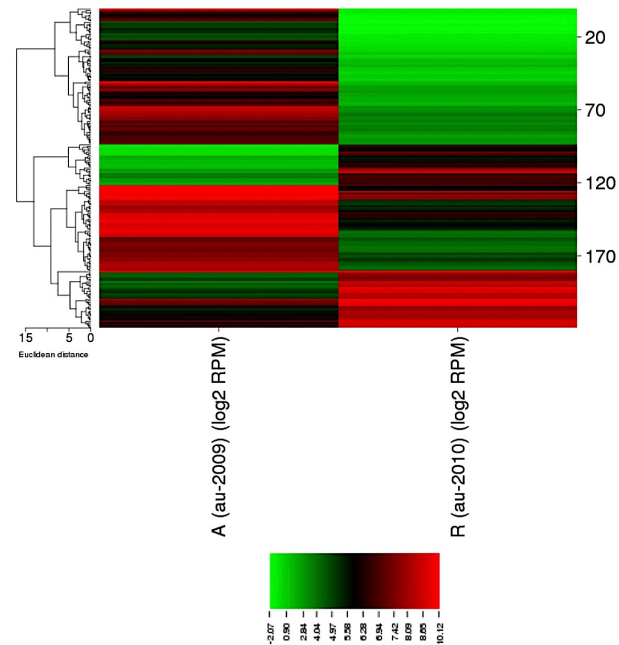
Two-component regulatory system

M00516 SLN1-YPD1-SSK1/SKN7 (osmosensing) two-component regulatory system [PATH:map02020 map04011] [BR:ko02022] (7)	complete
Drug resistance	
M00628 beta-Lactam resistance, AmpC system [PATH:map01501] (3)	incomplete
Cell signaling	
M00687 MAPK (ERK1/2) signaling [PATH:map04010 map04013 map04011] (1)	2 blocks missing
M00688 MAPK (JNK) signaling [PATH:map04010 map04013] (2)	2 blocks missing
M00689 MAPK (p38) signaling [PATH:map04010 map04013] (3)	1 block missing
M00677 Wnt signaling [PATH:map04310] (1)	incomplete
M00682 Notch signaling [PATH:map04330] (1)	incomplete
M00683 Hippo signaling [PATH:map04390 map04391] (4)	incomplete
M00695 cAMP signaling [PATH:map04024] (5)	complete
M00694 cGMP signaling [PATH:map04022] (1)	2 blocks missing
M00691 DNA damage-induced cell cycle checkpoints [PATH:map04110] (6)	complete
M00692 Cell cycle - G1/S transition [PATH:map04110 map04111] (2)	1 block missing
M00693 Cell cycle - G2/M transition [PATH:map04110 map04111] (6)	complete
Signature module	
Gene set	
Metabolic capacity	
M00615 Nitrate assimilation (1)	complete

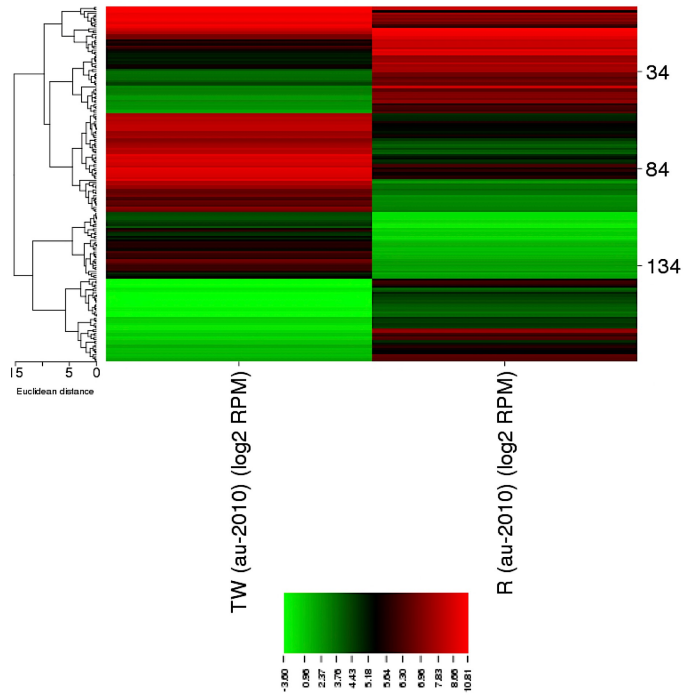
A (au-09) vs TW (au-09)



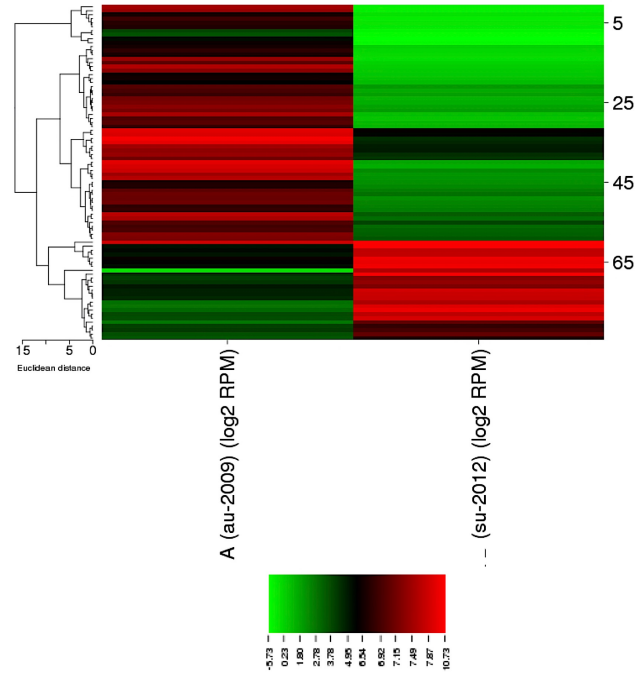
A (au-09) vs R (au-10)



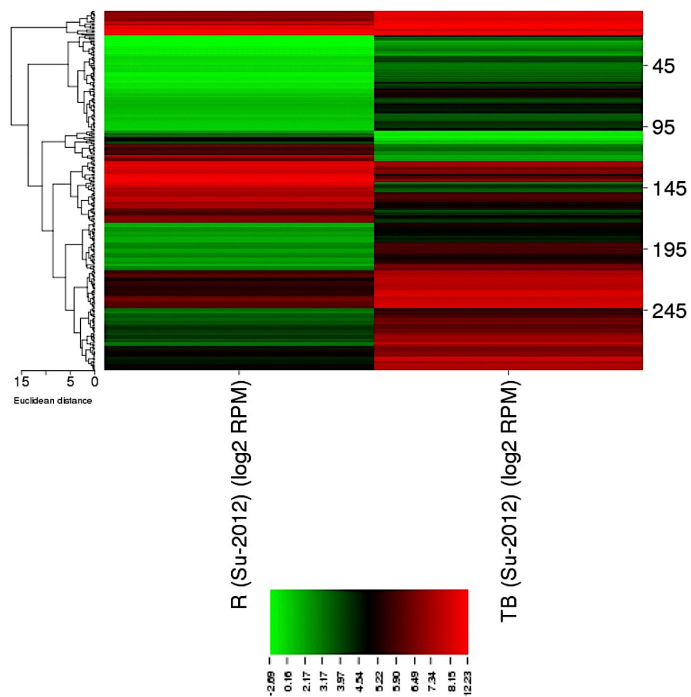
R (au-10) vs TW (au-10)



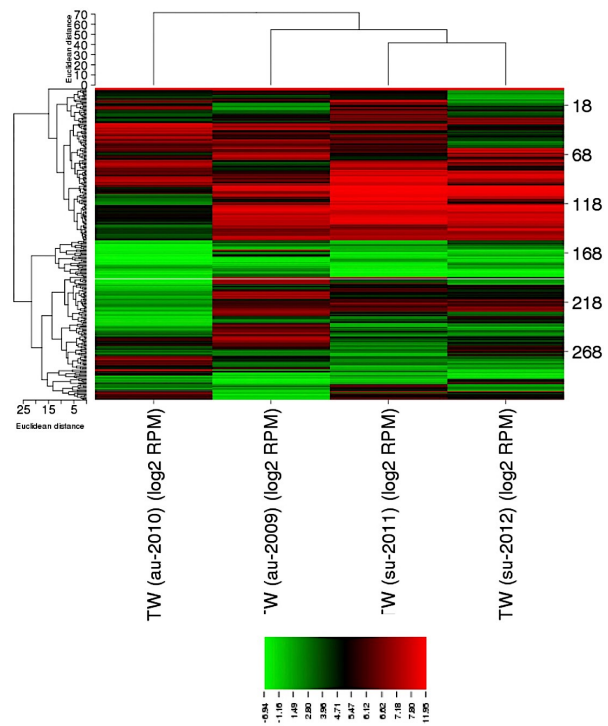
A (au-09) vs TB (su-12)



R (su-12) vs TB (su-12)



TW (au-09) vs TW (au-10) vs TW (su-11) vs TW (su-12)



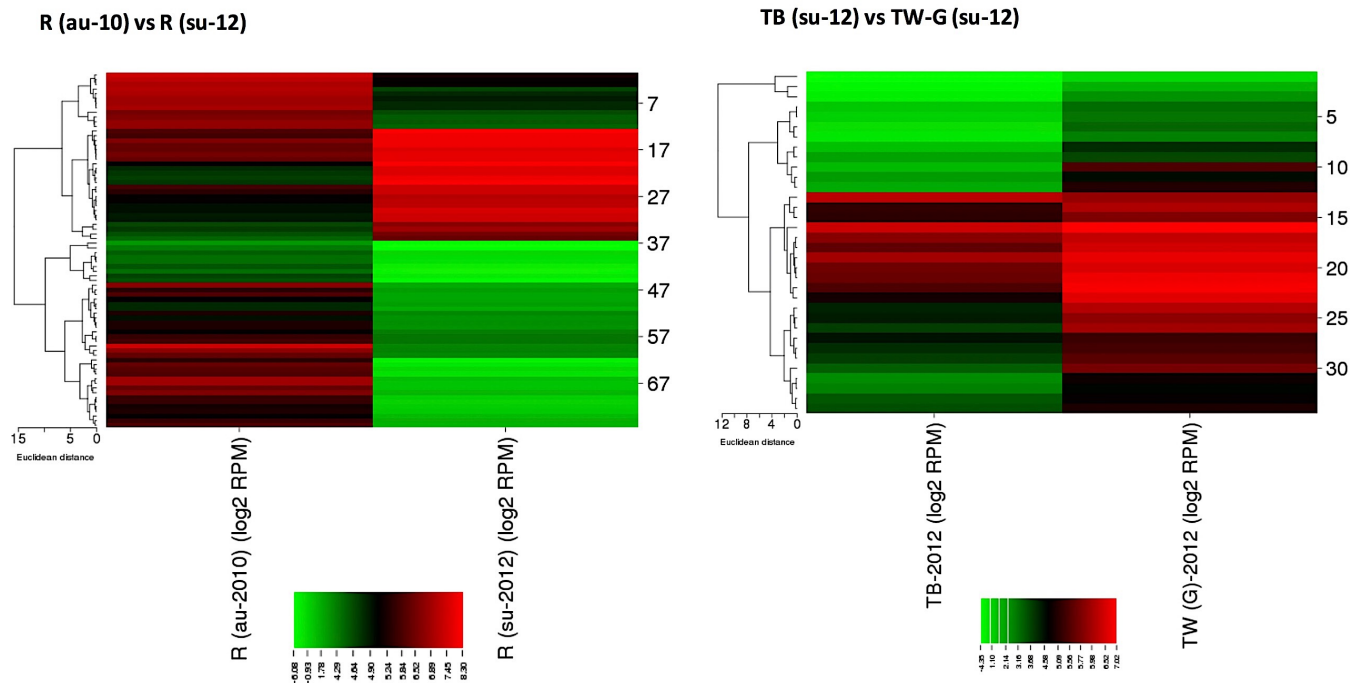
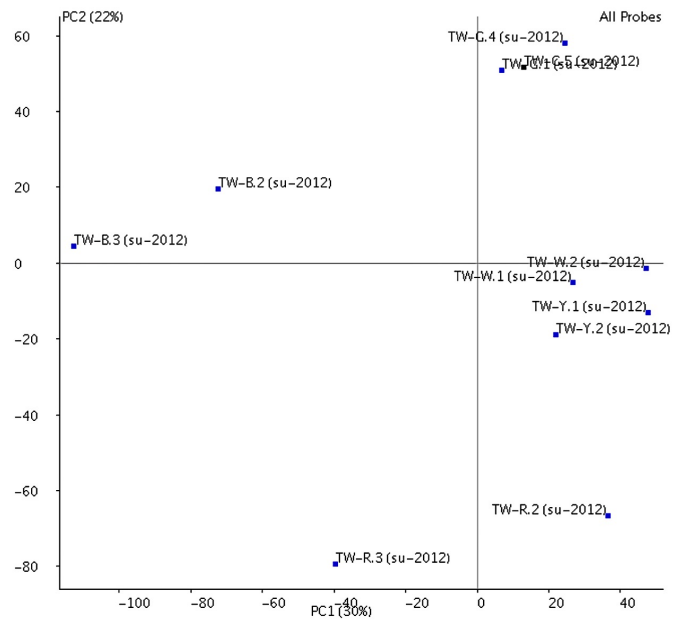
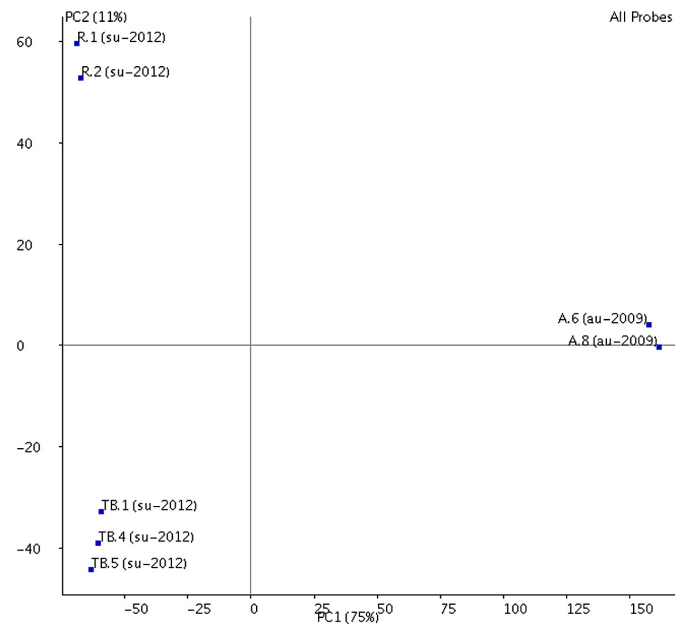


Figure S21: Heatmaps for hierarchical clustering of DEGs for the different tissue comparisons. Rows correspond to genes and columns to samples. Green and red colors indicate high and low expression, respectively.

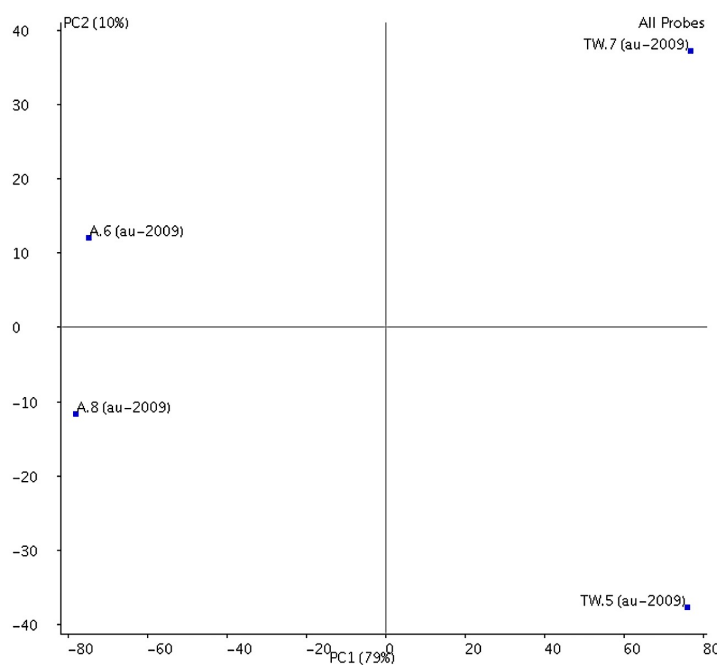
(i) TW-B vs TW-G vs TW-W vs TW-R and TW-Y (su 12)



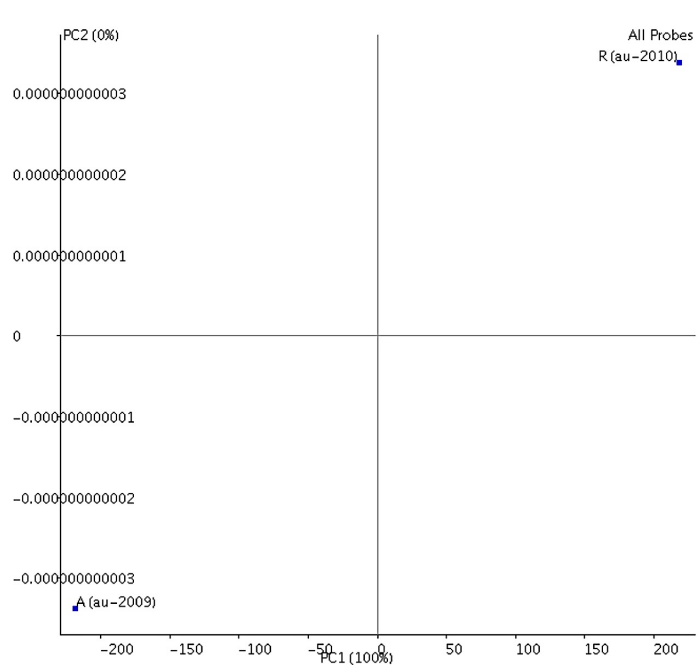
(ii) A (au-09) vs R (au-11 & su-12) vs TB (su-12)



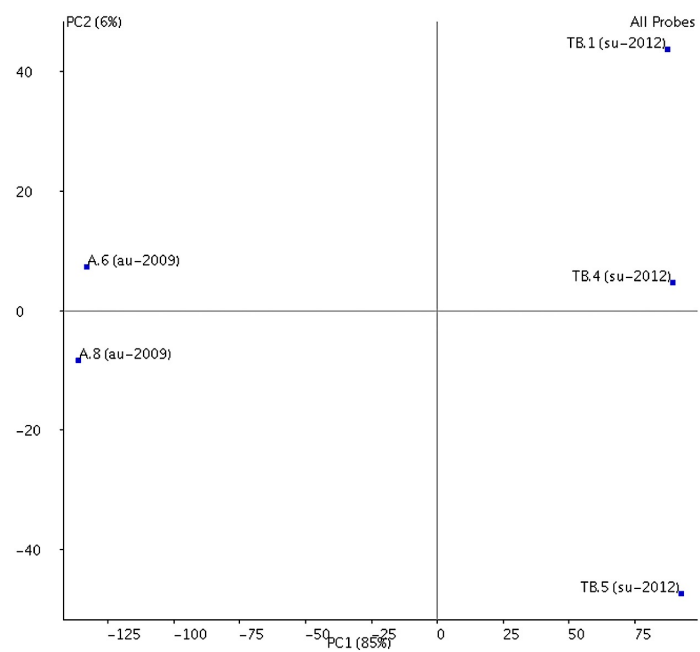
(iii) A (au-09) vs TW (au-09)



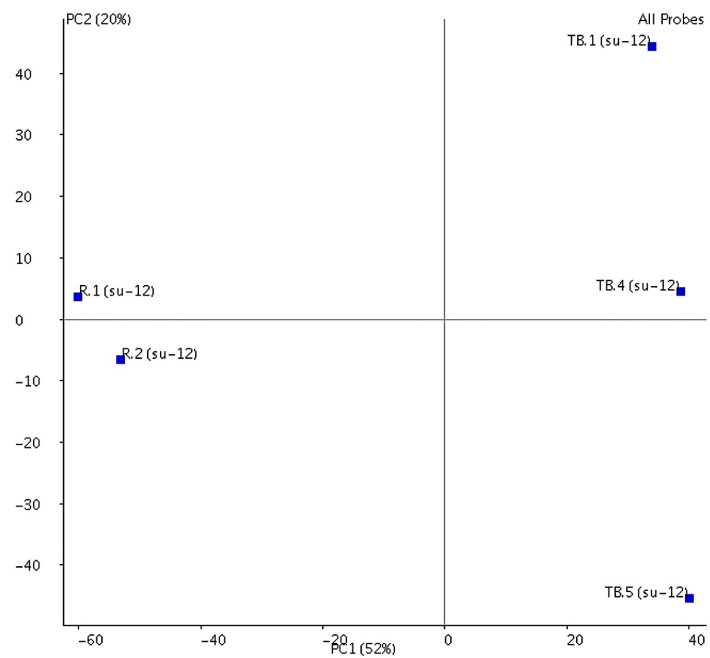
(iv) A (au-09) vs R (au-10)



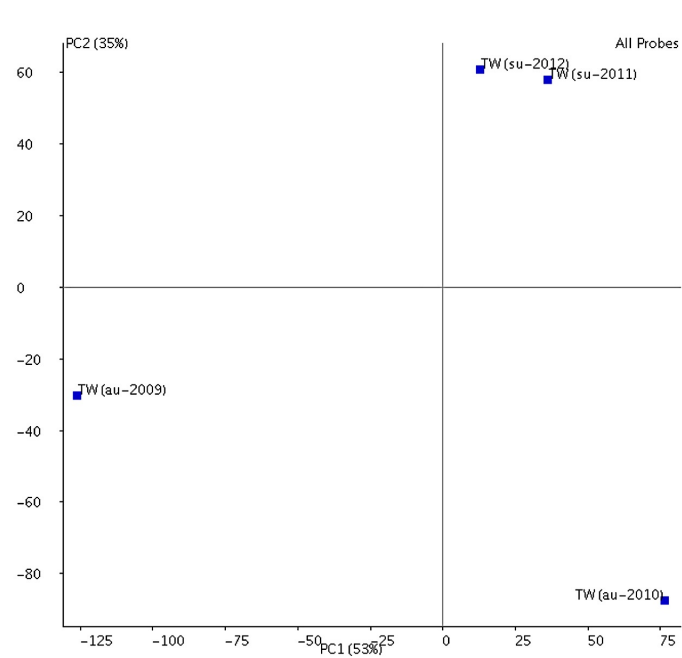
(vi) A (au-09) vs TB (su-12)



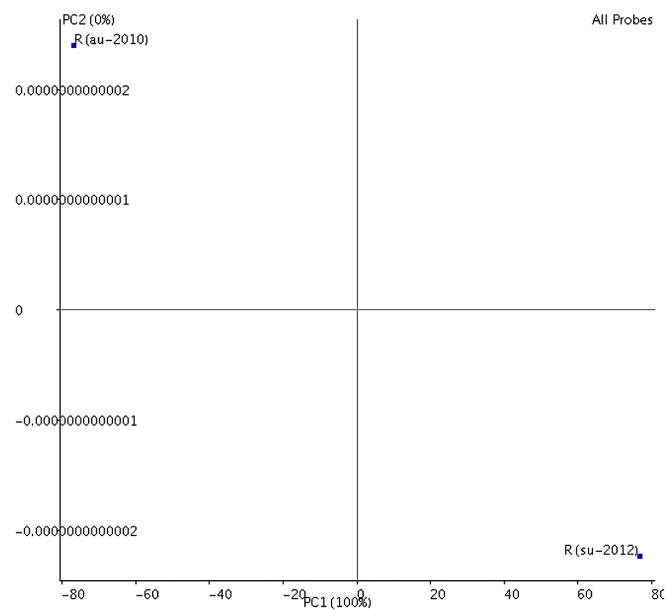
(vii) R (su-12) vs TB (su-12)



(viii) TW (au-09) vs TW (au-10) vs TW (su-11) vs TW (su-12)



(ix) R (au-10) vs R (su-12)



(x) TB (su-12) vs TW-G (su-12)

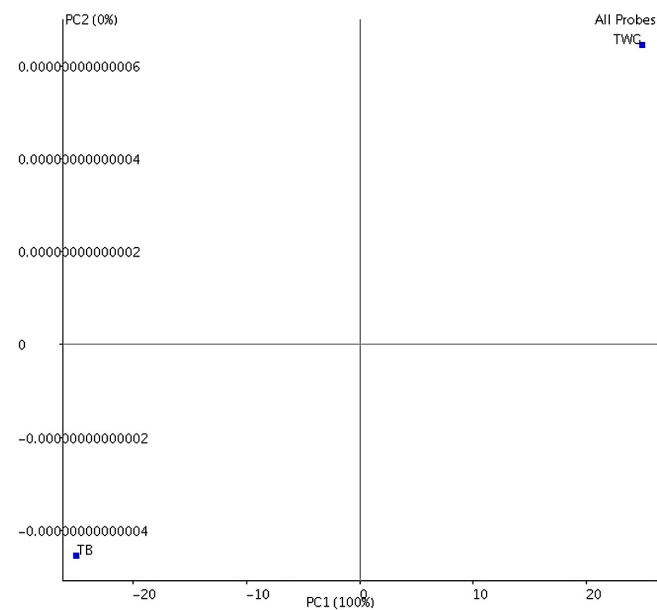


Figure S22: PCA plots of DEGs for the different tissue comparisons.

Appendix H: List of DEGs from all the tissue comparisons.

Table S18: List of 768 DEGs between TW–B vs. TW–G vs. TW–W vs. TW–R and TW–Y (su 12). The background variation genes i.e., genes with >2X fold change are highlighted in red.

Gene id.	Description	TWB (log ₂ RPM)	TWG (log ₂ RPM)	TWR (log ₂ RPM)	TWW (log ₂ RPM)	TWY (log ₂ RPM)
g53	thioredoxin	8.47	6.82	7.23	6.58	6.73
g101	hypothetical protein SBOR_10102	7.95	7.13	6.95	6.98	6.37
g110	Spore wall maturation DIT1	7.99	7.19	5.68	6.30	6.60
g111	pyoverdine dityrosine biosynthesis	7.86	7.03	5.47	6.14	6.53
g113	Saccharomyces cerevisiae YBR180W DTR1 dityrosine transporter	6.30	6.33	4.48	5.37	5.91
g114	related to alcohol zinc-containing	8.36	8.00	6.61	7.28	7.63
g144	hypothetical protein EMPG_15312	8.02	7.01	6.03	6.06	5.79
g174	30 kDa heat shock	6.20	5.86	4.55	5.21	4.71
g205	hypothetical protein DHEL01_12608, partial	6.05	6.23	7.40	6.39	6.49
g206	Clavamate synthase	5.56	5.72	4.80	5.63	4.58
g209	hypothetical protein EPUS_04454	8.28	8.32	7.79	7.99	7.63
g365	---NA---	8.04	6.29	6.63	6.10	6.03
g366	hypothetical protein Hypma_07800	8.70	7.10	7.51	6.87	6.83

g384	hypothetical protein Z518_10799	10.67	9.79	9.72	9.34	8.86
g394	---NA---	2.10	3.29	4.70	4.95	4.92
g395	homoserine kinase	2.27	3.95	5.33	5.32	5.35
g434	predicted protein	5.76	6.11	5.22	6.40	6.19
g450	HET-s domain	7.94	8.53	8.26	7.84	7.70
g499	Hydroxamate-type ferrichrome siderophore peptide synthetase	9.40	10.15	9.43	9.61	9.49
g616	related to spindle pole body	4.61	4.94	3.95	5.63	4.79
g670	hypothetical protein EPUS_01341	6.13	6.30	6.53	6.25	7.02
g695	related to cercosporin resistance	8.08	8.20	8.13	8.01	7.59
g768	GPI ethanolamine phosphate transferase 1	7.65	8.72	7.83	7.85	8.11
g777	hypothetical protein V490_00898	6.92	6.52	5.90	5.69	5.98
g821	meiotic recombinase Dmc1	6.19	6.86	7.14	6.08	6.78
g839	rna binding	8.00	6.36	6.65	6.33	6.33
g920	hypothetical protein L228DRAFT_270092	7.75	6.02	6.51	6.00	6.14
g938	Bet v1	8.15	6.65	7.09	6.67	6.73
g965	---NA---	4.42	5.37	4.84	6.15	5.45
g986	calcium-translocating P-type PMCA-type	8.21	8.16	7.38	7.66	7.06
g1056	MFS general substrate transporter	6.74	6.43	4.65	5.40	5.23
g1063	peptidase S41 family	8.44	7.78	7.49	6.94	6.82
g1098	hypothetical protein T310_6105	5.09	5.47	5.84	5.07	4.31
g1116	hypothetical protein O9G_004380	5.68	5.69	6.16	6.55	6.51
g1211	hypothetical protein CONPUDRAFT_165241	4.88	6.22	6.03	6.29	6.67

g1215	Serine ase 2	3.77	5.51	5.29	5.69	5.61
g1217	HAT domain-containing	4.39	5.99	5.98	5.94	6.27
g1315	hypothetical protein O988_04557	7.26	7.68	-2.53	7.47	7.47
g1316	-ribulosamine 3- chloroplastic	-0.64	-1.27	-2.53	-2.19	7.35
g1317	aldo keto reductase	5.10	5.14	-2.53	5.05	7.81
g1318	hypothetical protein PV05_01060	6.06	6.96	-2.53	6.57	6.80
g1319	hypothetical protein MGYG_05382	6.17	6.45	-2.46	6.33	6.58
g1322	hypothetical protein ANI_1_2290184	6.36	8.04	7.27	7.85	7.56
g1496	toxin biosynthesis peroxidase,	7.62	7.36	7.78	7.03	7.61
g1498	exo-beta- -glucanase	6.05	5.79	5.93	5.16	4.81
g1664	DNA replication factor C subunit Rfc5	6.25	6.49	6.85	7.12	7.36
g1703	Eukaryotic translation initiation factor eIF-1	7.79	6.00	6.49	6.13	6.19
g1748	---NA---	5.35	4.94	6.39	5.82	5.65
g1773	unnamed protein product	4.47	4.95	3.33	3.85	3.24
g1803	DUF221 domain	7.41	7.52	7.38	7.18	6.61
g1900	SAM-dependent methyltransferase	3.99	4.74	5.02	5.80	4.61
g1902	nonribosomal peptide	4.64	5.71	6.17	7.65	6.34
g1903	related to non-ribosomal peptide synthetase	0.73	2.22	2.12	4.36	2.89
g1904	Aldolase-type TIM barrel	3.68	4.13	4.48	6.34	5.16
g1967	hypothetical protein L228DRAFT_26732	-0.72	0.12	4.31	0.10	-0.41
g2074	hypothetical protein AOQ84DRAFT_391047	5.75	5.79	4.66	5.20	6.20
g2075	conserved hypothetical protein	5.53	5.50	-2.53	5.21	5.42

g2129	reverse transcriptase,	7.25	7.05	6.98	6.21	6.68
g2251	hypothetical protein RO3G_12376	7.19	5.56	5.82	5.37	5.36
g2271	SNF2 family helicase ATPase	5.29	6.10	6.87	6.48	5.93
g2320	cytochrome p450	6.15	6.36	6.07	5.49	6.54
g2341	unknown, partial	8.25	9.00	8.19	8.98	8.68
g2497	conserved hypothetical protein	5.84	7.35	6.60	7.25	6.65
g2502	predicted protein	4.78	5.62	4.81	5.90	6.18
g2552	hypothetical protein T310_8290	6.13	6.24	-1.96	5.80	3.75
g2673	---NA---	6.79	6.68	5.79	6.79	6.20
g2754	alpha beta-hydrolase	3.13	2.94	2.79	4.62	3.90
g2788	---NA---	2.13	3.35	4.86	4.11	5.06
g2789	hypothetical protein ANO11243_044680	2.23	3.79	4.00	4.28	5.08
g2889	hypothetical protein PADG_00764	8.59	7.88	5.96	7.03	6.74
g2896	plasma membrane proteolipid 3	8.01	6.03	6.61	5.97	6.03
g2911	Calcium-activated chloride channel regulator 4	5.84	5.89	5.25	6.37	6.05
g3000	hypothetical protein EV44_g3593	7.47	7.90	8.12	8.19	7.54
g3172	hypothetical protein V500_00208	5.82	5.89	6.31	6.71	6.76
g3191	---NA---	1.23	-0.03	-1.29	-2.98	0.09
g3378	carbohydrate-binding module family 32	9.09	9.31	8.99	8.49	8.19
g3496	Fungal transcriptional regulatory	4.09	4.80	3.98	4.62	5.52
g3506	hypothetical protein EPUS_02122	7.50	8.40	8.39	7.92	7.22
g3668	hypothetical protein AOQ84DRAFT_378787	10.03	9.55	9.38	9.18	8.58

g3702	Tyrosinase Q92396	6.44	6.77	6.25	7.16	6.29
g3801	---NA---	7.53	6.29	6.19	5.86	6.24
g3842	hypothetical protein AOL_s00054g438	7.91	6.70	7.07	6.57	5.54
g3881	60S ribosomal L10-B	8.00	6.76	6.86	6.73	6.66
g3893	hypothetical protein GLAREA_08855	7.12	5.90	5.90	5.37	5.90
g3916	hypothetical protein K441DRAFT_652280	4.53	4.34	5.59	4.77	4.85
g4102	predicted protein	0.94	1.08	0.46	-2.11	1.20
g4125	---NA---	4.27	5.58	5.50	5.84	6.22
g4214	hypothetical protein AURDEDRAFT_158037	5.29	5.48	6.42	5.36	5.78
g4217	[Laccaria bicolor S238N-H82] XP_001878005	9.10	7.76	8.30	7.61	7.41
g4220	---NA---	4.12	3.85	4.60	5.27	4.00
g4273	---NA---	2.02	4.39	2.77	4.88	3.00
g4318	DNA excision repair Rad16	5.96	6.07	6.69	6.14	5.66
g4391	hypothetical protein K443DRAFT_13795	3.44	2.88	5.10	5.32	5.79
g4441	hypothetical protein K441DRAFT_673373	-0.64	-2.59	4.05	3.35	0.89
g4509	1,3,6,8-tetra-HN reductase	7.01	7.45	7.59	6.78	7.24
g4526	hypothetical protein K441DRAFT_186449	9.11	8.79	8.90	8.27	7.90
g4552	hypothetical protein V502_00990, partial	5.31	5.46	5.47	5.25	5.97
g4583	hypothetical protein ANO14919_140150	5.50	6.18	5.50	6.46	6.13
g4586	hypothetical protein EURHEDRAFT_376324	2.17	4.77	4.51	4.79	4.74
g4587	---NA---	5.37	5.52	3.78	5.51	5.74
g4588	---NA---	4.72	5.59	3.53	5.86	5.95

g4598	hypothetical protein TCE0_033f08981	3.42	4.76	2.19	4.17	5.01
g4607	hypothetical protein OIADMADRAFT_35369	4.78	5.41	4.52	4.34	5.51
g4617	hypothetical protein HMPREF1541_06712	6.47	6.93	5.90	6.32	6.23
g4647	---NA---	4.34	3.31	3.10	4.09	1.56
g4651	PA14 domain	2.00	3.95	2.74	6.97	1.45
g4652	toxin subunit	0.44	1.74	0.46	5.12	-0.91
g4653	toxin subunit , partial	3.54	6.12	3.45	8.83	4.06
g4667	thaumatin family	4.95	4.79	4.85	4.67	6.05
g4677	serine threonine- kinase ATR	5.54	6.71	7.07	6.85	7.20
g4708	hypothetical protein DAEQUUDRAFT_600045	6.37	6.29	6.86	6.98	6.90
g4731	---NA---	2.63	1.45	-2.46	-2.48	-2.94
g4739	Similar to hypothetical protein [Tuber melanosporum Mel28]; acc. no. XP_002840333	3.62	5.03	5.64	5.52	5.94
g4752	putative d-mandelate protein	6.56	6.71	7.18	6.11	6.50
g4793	uncharacterized protein PAC_07873	2.34	3.08	5.20	3.71	3.92
g4871	hypothetical protein sscl_11g082640	3.14	3.84	0.87	2.33	2.52
g4884	hypothetical protein ANI_1_1854104	9.18	9.11	9.68	8.99	9.52
g4922	hypothetical protein FVEG_12244	2.92	4.65	3.35	5.11	4.62
g4985	cytochrome c, partial	7.46	5.87	6.32	5.87	5.92
g5028	---NA---	4.02	4.52	-2.53	4.79	4.84
g5047	---NA---	3.31	3.68	-1.96	3.77	0.27
g5122	hypothetical protein HMPREF1120_05181	-0.64	-1.52	4.10	3.53	3.44

g5149	origin recognition complex subunit 1	4.18	4.74	3.15	3.49	2.90
g5151	hypothetical protein PVAR5_0091	6.12	6.71	5.91	5.99	4.35
g5167	hypothetical protein P875_00075833	0.44	-1.10	3.97	3.83	3.56
g5168	Subtilisin-like protease	1.23	1.47	6.55	6.39	6.55
g5192	uncharacterized protein PAC_01761	7.20	7.33	5.78	7.52	7.36
g5410	hypothetical protein VE04_07924	7.53	8.17	8.72	7.72	8.30
g5491	Sac2 family	5.37	5.42	0.38	5.22	4.86
g5522	---NA---	1.34	-2.87	-2.46	-1.24	-1.46
g5562	hypothetical protein SPPG_00479	3.27	4.77	4.46	5.42	5.53
g5695	---NA---	2.88	3.23	5.04	5.13	5.21
g5785	prostaglandin G H synthase 2 cyclooxygenase pgh2 cox2	2.41	4.58	4.40	5.36	5.08
g5804	tyrosinase	4.34	4.70	3.26	3.67	4.86
g5818	Uncharacterized acyltransferase	6.13	6.15	5.11	5.94	6.03
g5868	hypothetical protein K432DRAFT_446181	-0.72	-3.31	2.50	3.76	1.15
g5969	tyrosinase	5.30	5.92	4.74	5.48	5.89
g6041	unnamed protein product	6.34	7.07	6.34	7.39	7.02
g6082	---NA---	3.88	4.33	4.96	4.28	3.28
g6117	histone h3	8.26	6.73	7.01	6.80	6.73
g6364	nuclear GTPase SLIP-GC	6.99	7.60	6.45	8.09	7.63
g6386	hypothetical protein V490_07430	5.80	6.39	5.75	5.32	4.34
g6389	---NA---	3.06	2.65	-2.53	-2.48	3.31
g6390	hypothetical protein NEOLEDRAFT_1180297	1.39	1.31	-2.53	-2.90	2.53

g6418	WD domain	5.36	5.43	3.59	4.92	5.36
g6449	uncharacterized protein RCO7_06198	5.78	7.02	6.85	7.32	7.36
g6505	---NA---	2.67	5.05	5.01	5.00	5.28
g6513	predicted protein	5.68	5.94	7.89	6.32	6.09
g6558	membrane fusion mating FIG1	4.76	5.02	-2.53	5.45	5.43
g6577	tRNA (cytosine-5-)-methyltransferase NCL1	5.14	5.52	4.52	4.44	5.54
g6595	PQ-loop repeat-containing 2	7.47	7.14	6.86	6.80	6.42
g6599	hypothetical protein ASPWEDRAFT_28951	7.16	5.75	5.73	5.35	5.31
g6615	related to calcium-independent phospholipase A2	7.86	8.50	7.52	8.29	8.34
g6633	[Laccaria bicolor S238N-H82] XP_001878005	2.44	3.22	2.21	3.64	4.56
g6642	probable G2 mitotic-specific cyclin-B1	4.76	5.11	3.93	5.20	4.75
g6670	PLP-dependent transferase	6.58	6.46	4.44	6.22	4.33
g6711	carbohydrate-binding module family 13	7.60	6.35	7.31	6.65	5.72
g6750	---NA---	1.10	3.95	3.23	3.98	4.24
g6826	hypothetical protein CY34DRAFT_813561	6.01	4.03	3.16	2.95	2.81
g6896	---NA---	4.59	4.80	4.86	4.77	2.44
g6916	asparaginase family	7.40	7.71	7.65	8.35	7.91
g6998	spindle pole body-associated sad1	2.83	-2.40	-2.53	-3.05	-3.01
g7085	TPA_exp: Tyrosinase	0.73	4.19	4.62	3.34	4.89
g7123	MFS general substrate transporter	6.60	6.97	6.63	6.29	6.74
g7150	MOSC N-terminal	5.01	5.62	6.57	5.55	4.86
g7180	benzoate 4-monooxygenase cytochrome P450	5.92	6.44	7.44	6.52	6.38

g7181	Cytochrome P450 oxidoreductase	5.25	5.63	6.72	5.61	5.59
g7201	related to RCR2 Endoplasmic reticulum membrane	8.34	6.91	7.18	6.64	6.70
g7223	predicted protein	9.32	9.03	9.02	8.14	8.85
g7265	Collagen alpha-5(VI) chain	6.67	7.50	6.00	6.94	6.43
g7286	---NA---	2.31	2.33	3.36	3.60	4.31
g7385	U4 tri-snRNP-associated 1	0.65	-1.11	-1.96	2.20	0.43
g7420	D-3-phosphoglycerate dehydrogenase 1	5.43	5.49	5.59	6.07	6.51
g7451	hypothetical protein PAXINDRAFT_167264	4.18	4.66	5.79	5.51	6.05
g7495	hypothetical protein IQ06DRAFT_80159	7.84	6.62	6.67	6.24	5.87
g7529	related to cytochrome p450 7a1	2.63	4.66	4.60	3.46	1.66
g7530	hypothetical protein LY89DRAFT_743438	3.14	4.73	3.99	2.70	1.05
g7532	cytochrome P450 monooxygenase	6.49	6.46	6.55	6.25	4.84
g7545	kinesin light chain	7.15	6.70	7.95	8.30	8.41
g7546	hypothetical protein AOCH_000998	3.07	4.02	4.85	5.49	4.72
g7604	---NA---	6.97	6.60	7.62	6.38	6.90
g7608	hypothetical protein PV07_03359	5.48	5.61	4.72	4.21	5.53
g7609	hypothetical protein AOQ84DRAFT_355343	6.56	6.41	5.46	5.07	6.41
g7698	hypothetical protein ASPFODRAFT_215217	4.93	5.50	5.74	4.19	5.21
g7763	---NA---	3.74	3.98	2.59	1.77	2.29
g7778	Similar to Uncharacterized protein YEL023C; acc. no. P39992	4.57	4.43	2.16	4.05	4.18
g7853	integral membrane	4.65	5.31	2.75	3.81	3.22
g7881	hypothetical protein PV08_05142	8.99	7.43	7.61	7.26	6.91

g7935	polyketide synthase	4.16	4.51	4.76	4.99	2.43
g7937	reducing type I polyketide synthase	4.87	5.25	5.36	4.83	3.02
g7982	hypothetical protein V490_00397	2.67	3.00	-0.50	3.33	3.77
g8015	alcohol dehydrogenase	10.54	9.50	8.39	8.83	9.03
g8016	HC-toxin efflux carrier TOXA	7.64	7.98	7.30	7.05	6.87
g8024	hypothetical protein ASPACDRAFT_54183	-0.72	1.12	1.37	-3.05	1.96
g8030	factor for adipocyte	5.18	5.83	-0.16	4.36	6.05
g8031	formylmethionine deformylase	4.73	5.06	3.29	3.75	5.23
g8107	hypothetical protein K432DRAFT_430842	1.73	3.58	5.54	5.07	5.79
g8119	hypothetical protein SAMD00023353_9400220	-0.64	3.88	4.21	4.18	3.89
g8160	Protein of unknown function	3.68	4.71	2.83	3.53	4.21
g8164	cytochrome P450	8.44	6.49	5.24	3.56	5.13
g8165	FAD-binding domain-containing	6.13	4.89	3.20	2.34	3.57
g8166	reducing type I polyketide synthase	3.69	3.57	0.71	1.05	1.65
g8168	reducing type I polyketide synthase	6.38	6.22	3.21	3.93	4.82
g8169	reducing type I polyketide synthase	5.35	4.81	2.57	3.20	3.58
g8182	WD40 repeat	4.15	5.19	4.50	4.36	3.16
g8184	ankyrin repeat-containing	6.48	7.46	7.71	7.39	8.04
g8276	NADPH dehydrogenase	8.73	8.32	8.61	7.27	8.31
g8290	salicylate hydroxylase	5.73	6.04	6.95	6.59	6.70
g8297	related to ankyrin 3	5.41	6.48	5.68	6.74	6.53
g8303	reverse transcriptase	4.86	4.70	1.60	2.85	3.82

g8313	integral membrane	6.14	6.22	6.23	5.28	5.66
g8383	hypothetical protein EPUS_03429	1.23	1.13	-0.66	0.33	-3.01
g8387	hypothetical protein EPUS_00229	3.00	3.89	1.56	3.88	3.95
g8406	dipeptidyl-peptidase III	4.81	6.49	5.75	5.26	5.08
g8421	hypothetical protein A7U60_g2462	3.71	3.21	1.72	3.42	-0.70
g8440	nuclear membrane fusion	5.68	6.45	6.92	6.69	7.25
g8443	hypothetical protein CPAR2_303200	3.76	3.81	1.05	4.08	-0.57
g8473	---NA---	7.39	7.46	7.15	7.52	6.97
g8491	F-box and WD40 domain ,	9.49	9.32	9.05	9.02	8.60
g8530	---NA---	0.94	1.60	0.09	2.77	4.11
g8560	GRIP domain-containing	3.39	2.35	4.81	3.73	3.14
g8567	---NA---	0.86	3.95	2.97	3.81	4.45
g8582	predicted protein	1.89	1.63	4.11	4.65	3.44
g8583	conserved hypothetical protein	-0.06	-0.36	3.72	3.97	2.46
g8609	Son of sevenless homolog 2	1.60	-0.04	-2.53	0.88	1.13
g8611	hypothetical protein AUEXF2481DRAFT_89676	3.41	2.90	-2.53	3.07	2.91
g8612	probable SMC3-required for structural maintenance of chromosomes	5.27	4.97	-2.53	5.23	5.06
g8613	hypothetical protein MVEG_03385, partial	3.69	3.28	-2.53	3.43	3.59
g8614	---NA---	5.21	5.39	-1.66	5.25	5.25
g8615	hypothetical protein AN958_10085	6.82	6.73	-2.53	6.73	6.77
g8616	hypothetical protein S40285_08438	0.76	1.28	-2.53	0.34	0.49

g8666	UvrD-helicase-domain-containing protein	4.50	5.15	4.30	5.58	4.30
g8692	---NA---	2.36	4.79	4.23	5.07	4.14
g8747	uncharacterized protein CPUR_04129	-0.72	-3.31	4.61	4.27	3.42
g8748	reverse transcriptase	0.15	-3.26	3.03	2.84	2.53
g8749	hypothetical protein PENSUB_11874	-0.72	-2.92	5.19	4.35	4.60
g8750	---NA---	-0.72	-3.31	2.67	2.28	1.70
g8755	hypothetical protein Z518_00039	3.02	3.89	5.09	4.26	2.39
g8779	benzoate 4-monooxygenase cytochrome p450	1.44	0.67	-0.18	5.35	5.75
g8862	related to hsp70	6.98	7.26	7.79	7.09	7.80
g8864	uncharacterized protein PAC_09661	6.48	7.09	8.29	7.46	8.38
g8882	hypothetical protein OIADMADRAFT_105967	6.51	7.13	6.65	7.19	6.61
g8922	Nonribosomal Peptide Synthase (NRPS)	2.26	3.27	2.34	4.19	1.72
g8969	Ankyrin repeat-containing	6.71	7.81	7.33	7.33	7.99
g9013	peptidase S28	1.62	2.09	2.30	-3.05	-3.01
g9037	sporulation-specific N-formyltyrosine oxidase	5.94	5.33	4.30	4.62	4.81
g9039	tetracycline-efflux transporter	7.27	7.15	6.15	6.68	6.94
g9041	related to major facilitator (MFS1) transporter	7.67	8.03	6.40	6.91	7.33
g9082	cytochrome P450 monooxygenase	3.94	4.82	5.85	5.86	5.92
g9086	terpene cyclase	3.69	4.09	4.21	5.17	3.75
g9095	predicted protein	5.68	5.90	5.55	6.44	6.46
g9272	MFS transporter	3.94	5.14	3.61	5.20	4.72
g9309	signal recognition particle, SRP19 subunit	9.34	7.77	7.57	7.22	7.22

g9420	hypothetical protein ZTR_09586	6.07	4.84	4.19	4.52	4.60
g9441	CP2 transcription factor family	7.97	7.96	8.40	7.78	7.32
g9491	L-fucose permease	3.63	4.52	4.93	5.28	4.22
g9503	hypothetical protein EPUS_00568	4.39	2.19	2.27	1.43	-0.30
g9532	hypothetical protein GGTG_03837	2.02	4.57	4.60	4.53	5.33
g9548	hypothetical protein FE78DRAFT_72758	5.56	5.40	4.08	5.01	5.54
g9600	calcium-channel CCH1	5.57	6.01	5.15	5.68	5.45
g9601	calcium channel subunit Cch1	6.05	6.35	5.38	6.11	5.69
g9663	2-hydroxyacid dehydrogenase	5.24	5.71	6.18	5.04	5.73
g9683	3-hydroxyisobutyrate dehydrogenase	10.43	10.30	9.81	9.77	9.64
g9685	hypothetical protein CISG_07311	-0.72	-3.26	3.94	3.64	3.84
g9702	calcium calmodulin-dependent kinase	8.83	8.76	7.72	8.20	8.08
g9710	---NA---	4.43	4.79	6.18	4.25	4.56
g9875	hypothetical protein TESG_03390	8.05	6.88	5.52	6.01	5.66
g10007	Integrase recombinase xerD	5.01	5.12	5.08	5.15	3.63
g10033	hypothetical protein AOQ84DRAFT_271286, partial	1.89	2.76	6.18	3.10	3.66
g10077	hypothetical protein V502_09580	4.38	4.43	3.31	5.27	5.06
g10169	hypothetical protein T310_3265	6.01	6.46	7.38	6.09	7.18
g10171	ankyrin	6.34	6.92	8.01	6.36	7.86
g10172	eukaryotic peptide chain release factor GTP-binding subunit	5.50	5.94	0.44	3.11	4.39
g10201	NGG1p interacting factor 3	8.26	8.27	8.40	7.91	7.57
g10209	cytochrome P450	6.45	4.60	2.81	0.88	3.37

g10212	cytochrome P450	6.90	5.30	3.17	2.42	3.84
g10213	cytochrome P450	5.03	4.41	2.38	0.89	1.79
g10214	short chain dehydrogenase reductase	5.61	3.37	1.60	0.34	1.20
g10215	oxidoreductase	8.82	7.48	6.84	6.48	6.73
g10250	hypothetical protein EPUS_09042	7.47	7.74	8.63	8.45	8.45
g10306	reducing type I polyketide synthase 10	5.85	7.23	7.07	7.23	7.74
g10355	related to 6-hydroxy-D-nicotine oxidase	4.92	4.69	4.98	5.85	5.48
g10367	hypothetical protein MPH_10379	5.12	5.85	5.67	4.09	3.57
g10421	reducing type I polyketide synthase	7.50	7.79	4.93	6.38	5.49
g10423	hypothetical protein ANO14919_092000	-0.72	-3.26	-2.53	2.09	3.37
g10454	trehalose-6-phosphate hydrolase	3.74	4.53	5.58	7.24	7.68
g10458	hypothetical protein VE00_03365	6.20	6.18	6.16	7.67	8.08
g10489	Retrovirus-related Pol poly from transposon TNT 1-94	4.68	1.73	2.60	3.55	6.11
g10505	---NA---	0.60	1.34	-2.53	1.13	1.42
g10509	mhck ef2 kinase domain family	5.24	4.91	6.57	5.30	5.50
g10514	benzoate 4-monooxygenase cytochrome p450	3.19	4.12	5.37	3.85	2.98
g10559	hypothetical protein K432DRAFT_354452	4.77	5.77	6.39	5.85	5.95
g10566	hypothetical protein EPUS_00474	6.90	7.22	6.99	6.39	5.87
g10571	non-reducing type I polyketide synthase	5.80	6.63	5.48	5.82	4.86
g10674	uncharacterized protein Y057_5681	-0.64	1.61	2.21	2.64	-0.35
g10755	---NA---	1.40	3.88	2.93	1.79	-0.56
g10843	hypothetical protein EPUS_00568	4.49	5.30	4.44	4.62	3.37

g10859	[Laccaria bicolor S238N-H82] XP_001878005	-0.72	-2.59	2.35	-3.05	-1.78
g10862	[Laccaria bicolor S238N-H82] XP_001878005	-0.72	-1.96	5.79	1.85	-0.46
g10884	Uncharacterized protein AUREO_015790	6.28	6.86	6.17	6.05	6.31
g10896	cytochrome P450	4.87	5.25	6.86	5.32	5.14
g10898	efflux pump antibiotic resistance	2.10	2.90	4.87	3.06	1.81
g10970	hypothetical protein UCREL1_2371	5.82	5.63	-2.46	6.39	6.74
g11004	galectin-like protein	11.51	9.91	11.04	10.27	10.04
g11015	cytochrome P450	3.99	4.84	0.04	5.03	5.21
g11044	[Laccaria bicolor S238N-H82] XP_001878005	4.39	5.58	6.15	6.18	6.35
g11045	glutaredoxin Grx1	7.86	6.36	6.37	6.28	6.45
g11090	predicted protein	8.58	6.90	8.05	6.81	6.82
g11126	cytoskeleton assembly control SLA1p	2.84	3.27	2.90	5.03	4.69
g11127	hypothetical protein EPUS_02021	8.54	6.63	6.78	6.21	6.27
g11131	---NA---	4.54	5.92	6.08	6.11	6.10
g11151	clathrin coat assembly	7.41	7.54	7.65	8.36	7.81
g11157	bacteriodes thetaiotaomicron symbiotic chitinase	3.52	5.80	5.39	5.52	6.05
g11170	related to lipoxygenase 1	4.97	5.22	5.47	5.63	4.20
g11171	Lipoxygenase	5.66	5.92	6.08	6.15	4.67
g11172	uncharacterized protein RAG0_16001	2.23	3.42	4.03	3.91	2.15
g11199	hypothetical protein T310_7253	0.60	1.53	0.41	3.56	1.99
g11214	hypothetical protein MMYC01_202431	5.31	5.61	6.88	5.26	7.00
g11319	uncharacterized protein PAC_15529	3.29	3.29	5.09	5.48	4.98

g11320	uncharacterized protein PAC_15527	5.82	6.09	6.84	7.10	6.93
g11321	hypothetical protein ACJ73_05116	7.60	6.91	5.53	6.00	6.27
g11334	Linear gramicidin synthase subunit D	7.20	7.52	7.07	6.74	6.64
g11335	conserved hypothetical protein	6.85	6.96	6.56	6.21	5.84
g11396	---NA---	4.69	4.24	4.62	3.70	1.25
g11398	briggsae CBR-LAT-2	4.15	4.36	4.36	4.13	1.84
g11400	---NA---	4.71	4.63	5.05	4.40	2.36
g11401	---NA---	6.30	6.28	6.74	5.89	3.80
g11402	hypothetical protein PICMEDRAFT_75192	6.52	6.09	6.57	5.85	3.89
g11406	NB-ARC and TPR domain (JCVI)	3.53	4.84	6.12	6.03	3.16
g11417	hypothetical protein K457DRAFT_101856	1.89	3.14	1.16	4.40	3.99
g11435	uncharacterized protein PAC_10069	4.95	5.62	4.71	5.10	6.26
g11436	related to multidrug resistant	2.61	3.64	3.07	3.07	4.68
g11437	uncharacterized protein PAC_17077	2.96	3.46	2.33	3.38	4.16
g11449	S-adenosyl-L-methionine-dependent methyltransferase	7.11	7.16	7.84	7.07	6.87
g11467	putative transposase	3.57	4.32	2.09	4.66	4.73
g11483	cytochrome p450	6.58	6.77	5.44	5.54	5.97
g11486	cytochrome p450	5.01	5.73	5.06	4.18	5.09
g11523	hypothetical protein ASPGLDRAFT_33331	2.50	0.45	5.66	4.82	5.39
g11596	glucose repressible grg1	8.75	6.81	7.41	6.80	6.97
g11622	hypothetical protein V498_10729	4.24	4.16	4.08	6.14	5.19
g11642	hypothetical protein MELLADRAFT_78386	7.11	7.22	6.10	7.07	7.02

g11668	alpha beta-hydrolase	6.11	6.27	6.51	6.33	5.58
g11670	hypothetical protein SBOR_6919	3.85	4.34	4.86	3.87	1.84
g11713	hypothetical protein M437DRAFT_59426	6.99	6.63	5.40	5.65	5.57
g11784	carbohydrate-binding module family 13	-0.72	-3.31	-2.53	7.66	7.63
g11807	predicted protein	5.83	6.67	5.95	7.01	6.70
g11908	C2H2 finger domain	7.33	7.17	6.59	6.42	6.19
g11923	hypothetical protein ANO11243_060950	8.85	7.34	7.00	6.26	7.15
g11940	---NA---	5.09	5.93	3.69	5.61	4.31
g11948	alkaline ase	2.81	2.44	4.82	3.77	4.99
g11970	Vegetative incompatibility HET-E-1	4.05	5.28	4.21	4.60	3.75
g11995	carboxypeptidase cpdS	5.78	5.31	3.87	3.44	4.98
g12011	---NA---	8.38	7.78	8.07	7.22	7.95
g12020	hypothetical protein LY89DRAFT_747417	8.73	8.22	7.73	7.74	7.34
g12032	membrane dipeptidase	5.28	5.65	6.18	4.35	6.10
g12033	gliotoxin biosynthesis	5.01	5.28	5.54	4.02	5.22
g12036	fungal zn(2)-Cys(6) binuclear cluster domain-containing	6.05	6.28	6.45	5.34	6.39
g12037	hypothetical protein VE00_10398	4.96	5.08	5.33	3.68	5.09
g12046	hypothetical protein BATDEDRAFT_24227	2.89	3.94	4.16	4.98	3.95
g12077	---NA---	3.33	5.40	4.93	5.51	5.35
g12095	Aristolochene synthase	7.40	6.55	5.60	5.81	5.96
g12122	---NA---	2.00	4.17	3.17	-1.19	2.62
g12126	---NA---	4.87	5.65	6.14	4.63	6.03

g12137	hypothetical protein K441DRAFT_652457	8.16	8.26	8.15	9.08	8.98
g12138	histidinol-phosphate aminotransferase	7.62	7.97	7.07	8.41	8.39
g12147	hypothetical protein PAXRUDRAFT_805983	1.02	1.15	3.55	3.05	4.00
g12148	---NA---	3.29	2.72	5.87	5.59	6.61
g12149	predicted protein	3.50	4.24	6.47	5.70	7.13
g12157	uncharacterized protein PAC_15196	3.91	4.06	-1.96	3.39	3.91
g12168	gag polymerase env poly ,	6.60	6.51	4.90	4.81	6.14
g12196	hypothetical protein Moror_5945	0.76	1.42	-2.53	1.27	-0.70
g12198	hypothetical protein LLEC1_07357	3.22	3.43	-2.53	3.23	1.20
g12199	hypothetical protein WG66_11260	7.15	7.81	-0.96	7.54	5.60
g12200	Chitotriosidase-1	4.77	5.13	-1.96	5.62	3.48
g12201	hypothetical protein GALMADRAFT_237585	5.90	6.44	-2.53	6.22	4.56
g12202	hypothetical protein SI65_06794	2.02	3.63	4.07	1.96	-0.41
g12203	gag-pol poly	5.07	5.04	5.06	3.17	5.49
g12243	---NA---	3.58	2.94	6.15	1.26	3.23
g12245	P-loop containing nucleoside triphosphate hydrolase	3.31	3.03	4.09	5.21	5.57
g12261	hypothetical protein EPUS_07439	9.79	9.80	9.26	9.22	9.20
g12262	hypothetical protein LAESUDRAFT_91127	2.77	1.38	4.75	4.45	0.89
g12299	hypothetical protein LY89DRAFT_743438	1.73	2.20	2.25	3.38	4.70
g12302	sodium calcium exchanger	8.00	8.46	-1.46	7.36	8.37
g12303	kinesin-2 85 kda subunit	4.92	5.10	1.78	4.39	5.14
g12306	hypothetical protein M404DRAFT_122096	3.86	3.56	-2.53	-1.00	2.91

g12329	NFX1-type zinc finger-containing 1	4.98	5.93	5.50	5.99	6.60
g12345	hypothetical protein PV09_07038	5.40	5.58	3.84	4.43	4.57
g12346	HET domain	2.79	3.61	0.29	1.22	2.31
g12351	hypothetical protein PLICRDRAFT_40416	4.52	4.73	3.67	5.18	5.91
g12394	hypothetical protein W97_07005	3.31	3.68	3.34	3.06	4.96
g12420	hypothetical protein CGLO_08849	0.94	2.34	3.76	3.92	5.03
g12424	hypothetical protein EPUS_00568	-0.72	-1.10	3.55	3.58	3.51
g12425	capsular associated	-0.72	-3.26	3.14	2.22	2.74
g12432	---NA---	0.23	2.64	-2.53	2.29	-3.01
g12488	arsenical-resistance ACR3	2.02	4.34	4.47	3.89	4.27
g12502	hypothetical protein M437DRAFT_29481, partial	5.95	5.90	6.18	3.46	5.34
g12505	solid-state culture specific	3.74	4.89	5.01	0.83	4.78
g12514	hypothetical protein PISL3812_00598	5.34	6.55	6.01	5.50	5.26
g12577	---NA---	0.60	0.31	-2.53	-0.01	0.91
g12589	G U mismatch-specific uracil DNA glycosylase	2.42	1.86	-1.16	3.99	-0.96
g12610	hypothetical protein FE78DRAFT_479396	0.65	-3.31	-2.46	-2.98	-2.94
g12619	hypothetical protein M433DRAFT_74313	2.44	2.33	-2.53	2.25	-2.94
g12620	---NA---	4.47	5.51	0.64	6.09	0.63
g12621	---NA---	1.23	1.29	-0.96	2.87	-1.86
g12624	hypothetical protein IQ07DRAFT_603591	3.94	4.73	5.79	5.86	4.94
g12629	---NA---	3.89	4.70	6.85	5.58	6.56
g12664	ABC a-pheromone efflux pump	7.82	8.21	7.56	7.87	7.85

g12672	conserved glycine-rich	8.02	6.95	7.39	6.49	6.36
g12738	---NA---	2.39	2.98	-2.53	2.88	2.58
g12739	---NA---	8.76	8.53	-1.66	8.55	8.58
g12741	hypothetical protein IQ07DRAFT_553401	4.66	5.17	-1.46	5.07	5.39
g12743	hypothetical protein T310_3447	4.82	4.55	-2.53	4.64	4.80
g12744	hypothetical protein HIM_10392	6.93	6.45	-2.53	6.54	6.74
g12753	hypothetical protein L228DRAFT_281926	2.88	3.57	3.97	2.66	0.59
g12774	gag polymerase env poly	4.00	4.31	-1.16	3.63	4.01
g12829	---NA---	6.61	4.85	5.28	4.82	4.93
g12843	hypothetical protein K432DRAFT_429412, partial	3.37	4.35	4.72	5.57	5.53
g12856	hypothetical protein SAMD00023353_9400220	5.93	5.48	4.31	4.33	4.84
g12860	predicted protein	0.73	1.77	4.34	4.85	3.80
g12902	related to dityrosine transporter	6.86	6.51	5.53	5.28	5.25
g12950	hypothetical protein V491_06394	5.65	6.20	5.67	5.07	4.87
g12960	uncharacterized protein Y057_4031	0.60	1.99	4.51	4.11	3.56
g12988	hypothetical protein V496_01796	6.86	6.78	6.65	6.25	4.32
g12999	integral membrane	4.52	6.99	5.33	6.36	5.99
g13007	hypothetical protein LAESUDRAFT_814441	2.10	2.63	1.85	-3.05	-3.01
g13008	cell wall biogenesis	1.34	0.04	-0.79	-3.05	-3.01
g13009	isoform Er16 of ankyrin-1	3.25	3.25	3.47	-3.05	-3.01
g13010	serine threonine kinase	4.55	5.38	5.37	-3.05	-3.01
g13011	serine threonine kinase	2.75	3.72	3.22	-3.05	-3.01

g13012	hypothetical protein EPUS_03941	6.67	6.80	6.54	-3.05	-3.01
g13013	Zinc finger CCHC-type	4.20	4.01	4.05	1.68	1.63
g13060	laccase-like multicopper oxidase	10.79	10.50	10.36	9.99	10.65
g13113	glycoside hydrolase family 16	8.45	8.51	7.47	7.93	8.11
g13121	serine threonine- kinase chk1	4.29	4.66	2.83	4.07	4.56
g13137	Cytochrome P450	3.73	4.56	5.32	4.64	3.70
g13149	Serine threonine- kinase hal4	-0.72	2.57	4.16	4.63	3.66
g13150	hypothetical protein ASPCAL07495	4.07	3.51	3.67	6.86	6.27
g13153	hypothetical protein LY89DRAFT_783720	5.73	5.55	5.27	4.38	5.70
g13174	uncharacterized protein FPRO_04301	5.84	6.15	5.33	4.95	5.04
g13186	kelch repeat	2.72	2.98	3.93	2.51	4.61
g13187	---NA---	0.76	-2.34	-1.59	-2.98	-0.41
g13197	hypothetical protein W97_05036	0.94	3.89	4.69	4.57	5.51
g13231	hypothetical protein COCCADRAFT_95513	6.75	7.41	8.08	7.37	7.18
g13250	Fe2 Zn2 regulated transporter	6.20	6.35	7.14	6.56	6.60
g13257	catalytic	8.32	7.96	7.44	7.11	6.96
g13288	nuclear condensin complex subunit Smc2	6.14	6.85	6.74	7.10	6.52
g13337	Calcium calmodulin-dependent kinase type	7.73	8.06	8.86	7.56	8.43
g13350	hypothetical protein SAMD00023353_9400220	4.15	4.81	4.63	4.13	1.78
g13377	MFS transporter	6.87	7.20	6.79	8.02	7.35
g13379	hypothetical protein ACN42_g2717	4.18	4.41	-2.53	4.29	4.16
g13380	---NA---	3.63	4.16	-1.96	4.23	4.03

g13381	predicted protein	3.52	3.22	-2.46	3.09	3.12
g13382	---NA---	4.73	4.84	-0.73	4.64	4.52
g13383	hypothetical protein E5Q_05663	1.60	1.88	-2.53	1.78	2.01
g13391	---NA---	3.99	3.94	-2.53	3.84	3.92
g13392	---NA---	2.76	2.66	-2.53	2.47	2.41
g13393	hypothetical protein CIRG_03319	4.87	5.11	-2.53	4.94	5.17
g13403	hypothetical protein PENARI_c006G07035	4.13	3.61	-1.59	0.59	-0.30
g13436	related to bimE	4.85	5.73	6.60	5.80	6.32
g13486	hypothetical protein PV07_12346	5.81	6.77	6.32	5.66	6.70
g13487	beta-ig-h3 fasciclin	6.31	7.01	6.98	6.08	6.68
g13492	---NA---	0.94	1.26	-1.66	-0.01	3.26
g13494	---NA---	5.19	5.06	-1.88	-1.74	-1.70
g13502	long-chain fatty-acid- ligase	8.40	7.92	7.98	7.62	7.12
g13505	---NA---	3.93	0.35	-2.53	5.05	1.97
g13507	---NA---	-0.72	-2.92	3.01	2.66	-3.01
g13508	cell division control 42	-0.72	-1.10	8.89	8.32	-1.78
g13509	Pkinase-domain-containing protein	-0.72	-2.54	6.84	6.53	-2.14
g13510	sodium calcium exchanger	-0.72	-3.31	5.17	4.57	-3.01
g13544	het domain-containing	2.73	3.26	4.83	3.00	4.06
g13631	WD40 repeat	7.74	7.82	7.73	7.65	4.97
g13653	cytochrome P450 monooxygenase	7.05	7.31	7.97	8.09	7.82
g13699	hypothetical protein AK830	8.53	8.63	-0.18	8.60	8.70

g13700	hypothetical protein HIM_11596	2.67	3.57	-2.53	3.79	3.93
g13701	---NA---	3.14	2.74	-2.53	2.21	2.67
g13702	---NA---	3.39	3.54	-2.53	3.20	3.38
g13704	---NA---	4.23	4.57	-2.53	4.14	4.41
g13723	hypothetical protein SISNIDRAFT_484147	5.48	5.20	4.68	4.77	3.73
g13740	hypothetical protein CHGG_01848	0.36	-2.59	3.25	3.23	4.18
g13757	TPR domain-containing	0.36	-2.92	4.94	4.39	1.00
g13758	uncharacterized protein CPUR_00784	2.67	-2.20	6.16	6.11	5.74
g13818	Aspartate--tRNA ligase, cytoplasmic	2.84	4.62	7.21	6.98	4.71
g13826	ankyrin	-0.72	-3.31	4.34	4.01	4.09
g13827	hypothetical protein K441DRAFT_668854	-0.72	-3.31	4.02	3.76	4.00
g13830	hypothetical protein V492_04989	-0.14	1.10	6.54	7.17	6.14
g13831	---NA---	4.29	3.64	-1.16	1.90	-0.02
g13832	hypothetical protein PTT_16691	2.23	2.01	-2.53	-1.40	-2.14
g13841	High-affinity methionine permease P50276	6.42	7.29	7.55	5.12	6.38
g13856	MFS general substrate transporter	6.59	6.46	6.89	6.23	4.70
g13857	MFS general substrate transporter	6.34	6.43	6.83	6.01	4.32
g13892	hypothetical protein L228DRAFT_259842	7.57	7.52	8.81	7.62	8.68
g13917	hypothetical protein PIIN_06735	5.03	5.31	6.60	4.27	5.35
g13928	---NA---	4.84	5.20	5.55	3.26	5.87
g13929	hypothetical protein LAESUDRAFT_640535	2.02	3.34	2.78	1.24	3.12
g13964	hypothetical protein FPSE_01344	4.33	3.40	3.73	4.16	0.30

g14003	hypothetical protein V500_10202	2.52	1.88	-2.53	-1.40	1.72
g14004	uncharacterized protein RCO7_11213	3.58	3.93	-2.53	2.13	2.50
g14007	hypothetical protein AZE42_07563	6.12	5.88	-2.53	0.78	5.74
g14036	glycoside hydrolase family 3	8.98	8.58	8.01	7.95	7.53
g14037	hypothetical protein L228DRAFT_262392	7.19	7.88	6.54	6.77	7.65
g14046	hypothetical protein T310_2636	2.06	3.26	4.54	3.12	4.66
g14052	cytochrome P450	-0.64	-3.26	-2.53	3.00	3.43
g14053	cytochrome p450	3.00	0.01	-1.46	6.26	7.13
g14117	hypothetical protein K432DRAFT_274115, partial	2.94	3.41	6.29	2.28	2.33
g14129	alpha beta-hydrolase	-0.72	-2.21	4.30	1.95	0.72
g14133	hypothetical protein ASPWEDRAFT_28951	6.38	5.30	6.05	6.69	5.87
g14173	hypothetical protein SS1G_01102	2.63	3.59	2.15	3.50	4.39
g14186	TPR domain-containing	-0.72	-3.31	1.23	4.07	3.76
g14201	---NA---	2.10	2.25	5.05	4.30	4.94
g14228	TPR-like protein	6.28	6.53	5.98	8.14	7.36
g14302	hypothetical protein BK809_0005546, partial	1.79	4.45	2.70	4.37	2.86
g14320	hypothetical protein ALT_0985	9.36	9.37	8.17	8.08	8.83
	<i>Saccharomyces cerevisiae</i> YER014W HEM14					
g14328	Protoporphyrinogen oxidase, a mitochondrial enzyme that catalyzes the seventh step in the heme biosynthetic pathway	4.44	4.22	4.40	4.41	1.34
g14329	hypothetical protein W97_02516	5.41	5.15	5.24	5.30	1.67
g14352	hypothetical protein VE03_02072	7.68	6.09	5.92	5.58	5.62

g14381	hypothetical protein TBLA_0E01410	2.81	3.10	3.49	-3.05	2.79
g14382	hypothetical protein CHTT_0044650	6.56	6.63	6.86	-3.05	6.72
g14388	---NA---	0.60	-2.21	-1.29	-2.19	-1.09
g14412	S-adenosyl-L-methionine-dependent methyltransferase	5.31	6.00	6.16	6.51	6.85
g14420	uncharacterized protein PAC_01092	2.56	3.15	0.85	4.60	1.79
g14423	uncharacterized protein PAC_01090	4.17	3.62	2.13	5.04	1.90
g14424	High-affinity methionine permease P50276	5.45	6.09	6.02	4.69	5.98
g14425	---NA---	0.76	2.13	-2.53	2.54	0.59
g14427	hypothetical protein NCAS_0B04140	0.94	1.85	-2.53	1.89	-0.01
g14431	mating-type MAT1-2-1	0.44	2.35	-1.96	0.43	2.07
g14432	hypothetical protein ASPACDRAFT_51623	7.32	6.44	-2.53	5.89	6.41
g14433	---NA---	3.39	3.76	-2.53	2.92	3.78
g14458	predicted protein	2.89	2.74	-2.46	1.90	1.11
g14467	hypothetical protein MBM_00672	4.55	4.34	4.65	5.16	4.25
g14469	hypothetical protein RMatCC62417_11442	0.84	1.26	5.36	0.00	4.39
g14470	hypothetical protein A4X13_g7656	3.29	1.80	6.36	-0.50	5.30
g14477	DnaJ-domain-containing protein	4.78	5.11	4.74	5.15	6.23
g14485	hypothetical protein W97_04984	1.92	4.88	4.06	6.34	7.78
g14516	hypothetical protein VE03_08214	5.70	6.54	6.81	6.98	7.14
g14518	vegetative incompatibility HET-E-1	7.62	7.97	6.79	6.92	6.41
g14540	Lysozyme	5.35	6.93	5.89	6.93	6.56
g14543	reverse transcriptase	5.12	4.76	5.58	4.93	3.65

g14553	hypothetical protein S40288_09664	-0.72	-3.31	6.10	-3.05	-3.01
g14556	hypothetical protein AN8309.2	3.28	4.16	5.44	5.53	5.00
g14561	---NA---	2.89	3.66	3.01	3.63	1.30
g14562	---NA---	-0.72	-1.11	4.31	-1.61	4.66
g14566	DNA-binding SAP	-0.72	-3.31	3.52	1.98	2.31
g14614	---NA---	1.10	2.51	2.28	1.39	-3.01
g14629	hypothetical protein W97_08269	9.17	7.78	7.36	7.09	7.54
g14632	hypothetical protein V490_00340, partial	2.48	3.76	5.63	4.16	4.60
g14649	hypothetical protein W97_08945	1.23	2.21	-2.53	-3.05	-3.01
g14650	hypothetical protein V499_07880	5.22	5.49	-2.53	-3.05	-3.01
g14652	hypothetical protein ABOM_001483	6.28	6.70	6.27	6.25	2.65
g14658	hypothetical protein K441DRAFT_698916	7.28	7.18	7.57	6.94	7.49
g14670	hypothetical protein A109_07999	2.76	3.03	5.57	3.63	2.77
g14671	hypothetical protein HCBG_06035	4.05	4.68	5.78	6.70	5.77
g14706	---NA---	1.23	-2.15	5.36	-3.05	3.84
g14707	NAD(P)-binding protein	4.34	-1.71	7.43	-3.05	5.50
g14708	hypothetical protein CC84DRAFT_1195088	2.84	-2.20	5.70	-3.05	3.75
g14709	hypothetical protein H101_03049	2.79	-1.96	7.26	-2.98	5.42
g14714	kinase-like protein	1.39	2.94	4.79	-0.32	2.21
g14731	hypothetical protein BofuT4_P088950.1	6.64	7.13	6.88	6.50	7.14
g14736	hypothetical protein Hypma_11659	5.57	5.13	3.04	2.05	2.89

g14759	YN16_YEASTRecName: Full=Putative uncharacterized protein YNL276C	-0.72	-3.26	2.63	-3.05	2.67
g14762	hypothetical protein K441DRAFT_650523	7.40	6.60	6.88	5.90	5.93
g14766	hypothetical protein HCAG_06438	6.41	6.19	2.08	6.63	0.22
g14767	hypothetical protein BC1G_15283	1.81	2.86	-0.48	3.09	-3.01
g14776	sphingosine hydroxylase	0.94	-1.22	-1.05	3.26	3.95
g14778	hypothetical protein J132_02389	3.39	4.03	3.61	3.67	0.84
g14783	hypothetical protein VE04_08529	5.29	5.44	4.63	6.45	5.43
g14784	hypothetical protein JAAARDRAFT_61465	1.10	1.30	1.41	0.05	-2.86
g14785	hypothetical protein PV05_02168	3.63	3.48	3.46	2.53	-2.14
g14791	hypothetical protein ACJ73_07215	4.43	4.53	-1.66	4.05	3.96
g14792	LALA0S07e05578g1_1	5.05	4.87	-1.96	4.88	4.63
g14794	hypothetical protein FIBSPDRAFT_796861	3.27	3.54	-2.53	3.12	3.78
g14808	---NA---	1.39	-1.90	-2.53	5.43	3.71
g14809	hypothetical protein M378DRAFT_858718	2.92	-1.84	-2.53	7.23	5.55
g14810	gag polymerase env	-0.56	-3.31	-2.53	3.30	1.25
g14813	uncharacterized protein RSE6_06897	1.52	2.31	-2.46	-0.82	-2.36
g14896	hypothetical protein SS1G_06147	6.72	6.50	-2.38	-3.05	6.32
g14944	---NA---	4.68	5.12	-2.53	4.04	5.13
g14945	hypothetical protein EPUS_07439	9.33	9.62	0.17	7.02	7.54
g14951	hypothetical protein AOQ84DRAFT_306480, partial	0.52	0.90	2.30	-3.05	-2.94
g14953	hypothetical protein	1.34	1.29	1.10	-2.40	0.52

g14956	---NA---	5.05	5.21	5.45	6.40	5.89
g14958	---NA---	0.36	0.25	-0.03	4.15	0.70
g14959	---NA---	3.00	3.75	2.05	3.24	0.63
g14990	hypothetical protein OIADMADRAFT_178813	1.48	0.49	7.40	0.73	0.20
g14993	ankyrin	1.29	-2.59	7.64	7.98	6.06
g14998	---NA---	3.07	-2.15	6.19	5.93	4.89
g15000	hypothetical protein ST47_g8211	7.15	7.52	7.47	6.87	5.52
g15005	hypothetical protein AGABI2DRAFT_190340	0.23	2.38	-1.29	2.78	-3.01
g15006	zz type zinc finger domain	3.76	4.47	-0.73	5.05	-2.14
g15008	---NA---	-0.14	-2.68	-2.53	0.74	6.51
g15026	ZIP zinc iron transport family	4.47	4.94	5.95	4.68	5.39
g15043	GMC oxidoreductase,	6.45	6.77	-2.53	6.36	4.37
g15053	ankyrin	1.39	3.85	2.07	3.69	3.29
g15056	related to dis1-suppressing kinase dsk1	2.79	3.75	-2.53	3.26	4.15
g15060	hypothetical protein LEMA_P041030.1	5.40	5.75	-2.46	3.97	5.65
g15061	hypothetical protein AGABI2DRAFT_176739	7.35	7.66	7.96	6.72	8.04
g15067	predicted protein	7.42	6.14	5.66	5.78	5.87
g15068	hypothetical protein MVEG_09614	6.92	5.66	5.68	5.04	5.31
g15072	hypothetical protein PTTG_27154	6.06	6.13	4.73	4.22	4.61
g15088	---NA---	-0.64	-2.02	-0.48	-3.05	4.93
g15092	hypothetical protein PFICI_07720	5.57	6.42	5.69	5.42	3.13
g15105	hypothetical protein CPAG_00491	2.35	4.65	6.39	5.08	4.94

g15122	WD40 repeat	-0.72	-3.26	-2.53	5.93	-3.01
g15129	hypothetical protein V501_07551	4.50	4.22	-2.53	3.14	3.14
g15130	endoplasmic reticulum	2.92	2.47	-2.53	0.26	0.89
g15148	Cell polarity (Tea1)	6.49	6.57	-1.66	5.59	6.61
g15152	uncharacterized protein FMAN_15521	2.02	1.66	-2.46	0.41	0.39
g15156	probable chitin synthase D	4.02	4.40	4.10	2.98	1.90
g15168	pepsinogen C	-0.72	-3.26	2.94	2.80	3.87
g15172	uncharacterized protein CPUR_00784	2.00	2.17	0.28	5.88	5.36
g15176	hypothetical protein ABOM_011974	0.94	1.44	1.44	1.77	5.30
g15178	hypothetical protein PV05_09409	7.55	7.71	-0.96	7.02	5.55
g15190	hypothetical protein GLOINDRAFT_7345	-0.06	0.25	-1.09	4.19	4.98
g15192	---NA---	-0.64	-1.27	-2.46	3.37	3.90
g15194	Phospholipase C 3	0.36	-2.54	5.12	5.46	5.53
g15195	hypothetical protein HETIRDRAFT_449559	-0.64	-3.26	2.61	2.80	2.90
g15196	---NA---	-0.14	-2.82	2.98	3.30	3.38
g15197	hypothetical protein EPUS_06399	0.60	1.53	-2.53	0.89	1.89
g15202	reducing type I polyketide synthase	6.87	7.42	6.93	6.69	7.12
g15209	hypothetical protein VCUG_01512	-0.64	-1.03	4.28	3.57	0.81
g15223	hypothetical protein PV06_11452	3.81	3.88	-2.53	3.06	-3.01
g15256	hypothetical protein PV07_11093	3.73	4.14	3.35	4.12	2.17
g15259	---NA---	2.42	2.69	3.33	4.73	4.49
g15264	hypothetical protein M422DRAFT_50974	3.81	4.18	4.13	1.42	4.42

g15271	spindle pole body	5.19	5.93	6.81	6.88	6.87
g15272	hypothetical protein ASPVEDRAFT_41603, partial	-0.64	-2.92	-2.53	-3.05	6.13
g15282	---NA---	0.23	-3.21	-2.53	3.92	1.89
g15283	---NA---	-0.14	-2.68	-1.96	2.74	0.49
g15293	Saccharomyces cerevisiae YJR138W IML1 Subunit of the SEA (Seh1-associated) complex	-0.14	-3.26	-2.53	3.66	1.92
g15298	hypothetical protein AK830_g6710	2.60	2.68	-0.79	0.48	0.37
g15300	conserved hypothetical protein	0.86	1.36	3.10	3.02	4.88
g15303	hypothetical protein PV08_06397	2.39	2.75	5.08	-3.05	4.70
g15304	---NA---	0.44	1.09	4.66	-3.05	4.27
g15305	hypothetical protein DOTSEDRAFT_37712	0.76	-1.49	6.06	-2.48	5.86
g15321	Tetratricopeptide-like helical	-0.72	-3.26	-2.53	5.03	-3.01
g15324	retrovirus-related Pol poly from transposon TNT 1-94	1.10	0.94	0.87	3.73	0.70
g15326	hypothetical protein ARAM_007547	1.94	0.30	-2.53	-2.48	1.65
g15329	hypothetical protein ACJ73_07607	7.12	7.37	2.79	6.61	7.48
g15347	hypothetical protein UA08_03298	3.36	4.42	3.50	6.35	6.68
g15353	hypothetical protein LY89DRAFT_633746	3.36	2.64	-2.53	4.25	-3.01
g15357	WD40 repeat	-0.72	-3.31	-2.53	3.68	-3.01
g15360	---NA---	4.13	4.77	3.94	5.63	5.01
g15361	vegetative incompatibility HET-E-1	3.11	4.27	4.49	3.67	-3.01
g15372	hypothetical protein ACN38_g11345	1.76	2.46	-1.22	0.95	-2.94
g15384	hypothetical protein BK809_0001302	3.36	3.80	0.71	4.15	5.12

g15386	---NA---	1.63	0.90	-2.53	1.27	-0.30
g15388	---NA---	1.10	1.42	-2.53	1.50	-0.36
g15393	hypothetical protein K443DRAFT_578331	1.10	0.87	3.79	3.20	3.66
g15402	eukaryotic translation initiation factor 3 subunit M	6.60	6.42	-2.46	5.91	6.60
g15422	Cytochrome P450	1.02	-2.92	5.93	-3.05	4.14
g15423	cytochrome P450	3.08	-2.26	6.74	-3.05	5.04
g15427	---NA---	-0.64	-3.26	-2.38	-2.98	3.17
g15428	---NA---	-0.72	-2.68	-2.38	-2.40	3.52
g15432	hypothetical protein EPUS_06399	-0.72	-3.31	2.68	2.42	3.53
g15437	kinesin light chain	1.44	0.97	1.63	3.40	3.57
g15442	---NA---	5.92	5.64	6.76	6.49	6.52
g15453	uncharacterized protein FPRN_03580	-0.72	0.79	-1.29	3.30	0.68
g15458	hypothetical protein GALMADRAFT_225295	2.10	2.57	-0.29	4.23	1.66
g15465	TPA_exp: hypothetical protein A8136_1488	3.18	2.87	3.11	3.13	-3.01
g15478	hypothetical protein HMPREF1541_07223	6.52	5.17	4.96	4.77	3.62
g15482	---NA---	3.88	3.73	4.55	3.89	5.20
g15486	hypothetical protein PILCRDRAFT_526	0.60	-2.73	4.13	-3.05	2.28
g15487	kinase domain-containing	-0.72	-3.31	3.30	-3.05	1.63
g15501	hypothetical protein PFICI_15058	4.42	3.51	4.96	5.09	4.84
g15502	lectin-2	5.82	5.14	6.70	6.81	6.73
g15520	hypothetical protein LY89DRAFT_328989	1.34	0.43	-2.53	0.64	1.02
g15536	---NA---	2.08	2.42	-2.53	1.85	0.15

g15537	hypothetical protein EPUS_08934	-0.72	-3.31	-2.53	-2.98	5.66
g15544	hypothetical protein V492_07287	-0.72	-2.49	2.19	2.20	3.56
g15552	hypothetical protein NECHADRAFT_87699	1.34	2.10	-2.53	1.86	-0.86
g15557	hypothetical protein TOPH_06261	1.23	-2.59	-2.53	2.91	2.45
g15566	---NA---	2.39	1.71	-2.53	0.89	1.88
g15567	---NA---	5.49	5.25	-0.09	4.16	5.43
g15570	hypothetical protein PENSUB_4128	0.23	0.54	3.63	5.09	4.18
g15571	---NA---	1.63	0.89	0.35	4.04	3.02
g15585	serine threonine- kinase nrc-2	5.87	6.13	-2.53	5.33	6.00
g15586	BA75_05108T0	1.52	1.67	3.96	3.84	1.75
g15592	hypothetical protein	1.34	3.10	4.82	5.62	5.58
g15609	hypothetical protein G647_03093	1.89	3.72	5.74	4.52	5.24
g15625	hypothetical protein Z520_01840	4.20	3.98	4.97	3.53	5.33
g15626	hypothetical protein K432DRAFT_430842	-0.72	-3.31	2.38	1.29	3.13
g15631	hypothetical protein PISL3812_04397	-0.72	-2.73	6.55	5.07	6.49
g15637	hypothetical protein ASPBRDRAFT_139323, partial	4.29	3.64	4.88	2.26	3.97
g15641	---NA---	9.67	8.53	8.45	9.21	8.67
g15650	hypothetical protein HYDPIDRAFT_29501	-0.72	-2.34	3.92	-0.11	2.34
g15655	---NA---	-0.06	-0.77	-0.96	2.39	1.80
g15656	benzoate 4-monooxygenase cytochrome P450	4.52	-1.64	8.21	-2.19	6.46
g15666	---NA---	1.94	1.95	0.21	-2.40	-0.70
g15667	hypothetical protein MYCTH_99456	3.95	4.31	3.05	-1.40	2.34

g15682	uncharacterized protein PAC_18094	3.77	4.20	-2.53	3.67	-3.01
g15698	mfs drug efflux	4.14	4.74	5.72	2.28	1.61
g15700	hypothetical protein FPSE_08062	5.88	5.24	5.28	4.67	-3.01
g15701	hydrolase, partial	6.67	6.78	6.77	7.88	7.35
g15702	hypothetical protein I308_02452	1.62	2.81	5.31	3.60	3.78
g15713	hypothetical protein T310_3265	3.81	-1.56	7.32	7.07	5.21
g15743	hypothetical protein ARAM_003398	1.10	2.66	2.98	2.49	5.52
g15744	hypothetical protein BC1G_01702	7.37	7.17	6.80	6.42	4.95
g15745	hypothetical protein V496_01357, partial	1.60	1.95	-2.53	-3.05	-2.94
g15763	Retrovirus-related Pol poly from transposon TNT 1-94	2.23	1.94	3.40	2.65	5.04
g15767	---NA---	-0.72	-3.31	2.86	2.43	-3.01
g15780	---NA---	1.84	2.34	-2.53	-0.09	2.92
g15791	---NA---	-0.14	-1.82	3.66	2.98	0.56
g15797	gag polymerase env poly	3.16	3.10	-1.16	3.37	-3.01
g15800	verA monooxygenase	6.42	6.06	5.84	5.41	4.25
g15801	cytochrome P450 monooxygenase	1.60	-2.73	-2.53	3.21	3.53
g15804	---NA---	3.34	3.53	1.79	-3.05	-2.94
g15805	---NA---	4.88	4.51	3.91	-3.05	-3.01
g15809	---NA---	1.02	1.35	0.61	-3.05	1.34
g15826	hypothetical protein HYPsUDRAFT_209455	4.13	4.78	4.52	3.99	2.52
g15831	hypothetical protein VE03_06097	-0.72	0.66	3.39	2.61	4.01
g15855	Tip elongation aberrant 1	3.18	3.84	5.51	5.27	3.22

g15857	hypothetical protein CYLTODRAFT_443449	1.34	-1.05	-0.15	0.68	3.15
g15865	chromosome-associated kinesin KIF4B	6.18	6.46	8.01	7.88	6.09
g15866	hypothetical protein TCEO_015f02403	-0.06	-2.92	4.34	-3.05	1.87
g15876	CIC11C00000004402 [	2.77	2.90	4.93	4.71	3.65
g15899	conserved hypothetical protein	1.23	1.78	0.57	0.49	-3.01
g15923	hypothetical protein RMatCC62417_15089	0.84	0.78	-2.53	1.21	-2.44
g15924	hypothetical protein TOPH_08106, partial	3.55	3.46	-2.53	2.67	-3.01
g15925	uncharacterized protein PAC_02000	6.72	6.22	-2.53	5.82	-3.01
g15944	Mitochondrial chloroplast ribosomal L6	5.76	5.67	5.90	6.50	5.91
g15950	hypothetical protein TGAM01_05520	1.98	0.85	1.44	4.15	2.39
g15953	hypothetical protein L228DRAFT_237274	4.60	4.93	3.80	5.22	5.17
g16010	hypothetical protein EPUS_00568	7.13	6.72	7.77	7.95	7.41
g16020	hypothetical protein I303_05548	3.33	3.23	0.57	4.10	0.39
g16046	hypothetical protein ASPGLDRAFT_53618	3.34	4.21	3.34	-3.05	-3.01
g16064	hypothetical protein RO3G_04810	-0.72	-3.31	-2.53	-2.98	4.31
g16071	kinesin family	5.24	5.49	-0.15	4.73	5.47
g16106	hypothetical protein HMPREF1544_07045	-0.64	0.44	2.76	2.68	-0.36
g16111	---NA---	2.46	1.97	0.47	-2.98	-3.01
g16112	---NA---	3.63	3.54	2.46	-3.05	-1.78
g16118	---NA---	4.34	4.48	4.85	-3.05	2.57
g16148	hypothetical protein EPUS_00474	1.63	2.65	-2.53	2.23	3.18
g16152	Histone acetyltransferase complex SAGA ADA, subunit ADA2	3.46	3.04	-2.53	-1.61	0.96

g16168	hypothetical protein L228DRAFT_242344	-0.72	-3.21	-2.53	7.66	-0.46
g16179	---NA---	2.10	1.48	6.46	5.74	6.37
g16181	---NA---	-0.72	-2.68	3.65	3.23	1.42
g16184	---NA---	0.44	-2.59	4.89	4.65	4.70
g16189	hypothetical protein FOXB_14368	3.57	3.04	4.70	4.66	3.98
g16197	putative WD-repeat protein	5.52	5.87	2.99	5.90	6.23
g16201	hypothetical protein MGYG_05504	2.02	2.67	-2.53	2.02	-3.01
g16225	hypothetical protein SNOG_12454	4.04	3.52	2.00	5.20	2.80
g16226	hypothetical protein V501_09474, partial	0.94	1.21	2.35	1.57	3.82
g16228	---NA---	1.84	1.85	-1.66	-2.48	1.89
g16238	hypothetical protein PV04_07742	-0.72	-0.16	-0.88	3.54	-2.94
g16240	hypothetical protein AYO21_06735	-0.72	-3.21	-1.96	2.54	-2.14
g16248	hypothetical protein ASPCAL01128	3.13	3.30	4.15	4.08	-0.04
g16249	kinesin light chain	2.10	3.52	5.05	4.96	5.04
g16250	---NA---	5.08	4.86	-2.46	4.37	-3.01
g16251	hypothetical protein EPUS_03150	-0.72	-3.31	2.27	0.66	-0.54
g16252	Protein of unknown function	0.94	1.17	-1.29	0.18	-2.94
g16256	hypothetical protein T551_03738, partial	1.44	1.49	-2.53	0.69	3.03
g16259	hypothetical protein PMG11_10392	2.29	1.47	0.62	5.17	4.64
g16269	uncharacterized protein RAGO_12524	1.60	2.79	-2.53	-3.05	2.59
g16284	---NA---	3.86	2.89	1.80	5.26	-1.09
g16287	---NA---	6.37	6.00	5.15	5.52	5.40

g16299	hypothetical protein SPPG_00758	1.60	1.42	-2.53	-2.48	-2.94
g16302	---NA---	4.41	4.01	4.08	4.37	0.77
g16333	hypothetical protein H634G_07533	6.58	6.87	7.32	6.87	5.17
g16344	predicted protein	0.73	-0.10	2.79	2.06	4.14
g16349	hypothetical protein PAXINDRAFT_53669, partial	-0.72	-1.77	2.07	-3.05	1.48
g16354	hypothetical protein V502_09950	-0.72	-3.31	-2.53	5.11	4.08
g16358	hypothetical protein LY89DRAFT_739923	-0.72	-3.26	-2.53	-2.98	7.17
g16365	related to ribokinase	0.15	-3.26	5.19	5.47	3.48
g16372	hypothetical protein HMPREF1120_04477	3.70	4.27	-2.53	3.13	4.12
g16377	---NA---	0.15	-3.26	3.02	1.82	2.44
g16384	hypothetical protein PSTG_08909	1.60	1.65	2.37	1.78	4.59
g16385	hypothetical protein BC1G_04959	-0.72	-0.02	2.07	1.31	3.63
g16389	reverse transcriptase	0.44	0.89	0.82	3.11	4.04
g16399	---NA---	3.02	2.69	1.92	7.40	3.63

Table S19: List of 413 DEGs between A (au-09) vs. TB (su-12) vs. R (su-12), and the 74-background variation DEGs are highlighted in red.

Gene id.	Description	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
g43	hypothetical protein EPUS_07929	4.55	6.22	4.03
g144	hypothetical protein EMPG_15312	5.21	7.00	7.98
g199	hypothetical protein ZTR_09325	2.82	6.07	8.06
g238	catechol o-methyltransferase	4.94	6.50	4.16
g239	Dehydroquinase synthase	4.88	7.86	5.65
g246	---NA---	8.25	1.72	2.15
g271	class III chitinase 1	5.11	6.37	4.41
g272	glycoside hydrolase family 18 , partial hypothetical protein	4.72	6.38	4.60
g442	LY89DRAFT_716349	8.07	6.70	8.55
g489	---NA---	3.52	5.70	3.18
g505	geranylgeranyl pyrophosphate synthetase	7.67	7.05	5.65
g518	snoal-like domain-containing conidial pigment biosynthesis oxidase	4.79	6.50	7.68
g530	Abr1 brown 1	3.92	6.12	8.50
g534	aromatic peroxygenase	3.74	6.04	4.23
g548	Choline transport hypothetical protein	5.86	5.92	3.92
g578	AOQ84DRAFT_347259	7.48	7.91	6.63
g581	ABC transporter	7.04	2.77	2.69
g689	---NA---	7.16	1.53	1.38
g816	hypothetical protein TGAM01_04190	6.89	3.27	3.76
g967	glycosyltransferase family 2	4.05	5.20	2.56
g1001	udp-glucuronosyltransferase 2a3	6.39	7.11	4.58
g1005	flocculation suppression	4.07	6.81	4.59
g1006	hypothetical protein VF21_09233	4.14	7.32	4.12
g1044	vanillyl alcohol oxidase	3.88	6.35	3.12
g1045	amp-binding enzyme	4.18	6.90	3.83
g1047	amidase	2.01	0.44	-2.22
g1049	hypothetical protein	2.86	5.85	2.12

COCC4DRAFT_37502				
g1050	2og-fe oxygenase family oxidoreductase	2.78	7.09	3.93
g1051	hypothetical protein HIM_05906	2.09	6.68	3.01
g1052	ABC transporter	5.58	5.35	1.71
g1053	---NA---	4.56	5.26	1.88
g1125	zinc finger domain-containing	2.60	0.67	-1.75
g1155	polysaccharide deacetylase family	0.97	3.60	-2.08
g1213	hypothetical protein PIIN_08702	3.36	5.09	2.35
g1256	FAD dependent oxidoreductase superfamily	7.59	4.90	3.46
g1263	subtilisin-like protease	2.05	1.15	-3.13
g1316	-ribulosamine 3- chloroplastic	6.87	-2.14	-2.48
g1318	hypothetical protein PV05_01060	4.05	5.40	6.47
g1322	hypothetical protein ANI_1_2290184	-0.81	4.98	0.14
g1582	uncharacterized protein PAC_09565	4.99	4.96	6.34
g1602	related to FAD-binding	4.25	6.56	5.02
g1639	Piwi-domain-containing protein	6.94	0.67	-0.38
g1683	hypothetical protein RSAG8_11336, partial	-1.59	1.59	-2.21
g1734	hypothetical protein L804_04683	3.28	1.90	-2.59
g1816	murein transglycosylase	6.15	7.97	5.37
g2053	---NA---	8.62	4.69	4.41
g2086	related to Oxidoreductase, short-chain dehydrogenase	4.64	5.22	3.92
g2113	hypothetical protein W97_03127	7.62	0.81	1.50
g2498	hypothetical protein EPUS_03937	8.45	8.48	7.04
g2502	predicted protein	3.60	4.27	0.66
g2503	hypothetical protein EV44_g1971	5.03	5.36	1.06
g2504	hypothetical protein ASPZODRAFT_1392612	5.67	5.15	2.31
g2673	---NA---	7.18	6.31	4.95
g2698	A5178_ARTBC ame: Full=Uncharacterized secreted ARB_05178 Flags: Precursor hypothetical protein	4.88	3.94	6.29
g2749	BAUCODRAFT_27901	6.35	-0.25	-1.93
g2882	C6 transcription factor	7.16	5.21	3.21

g3057	Collagen alpha-5(VI) chain	5.12	6.82	7.47
g3168	amino acid transporter	6.60	6.72	5.27
g3172	hypothetical protein V500_00208	6.06	4.75	1.55
g3190	potassium uptake	5.94	5.49	7.28
g3282	hypothetical protein FOYG_00325	7.48	4.95	4.21
g3453	hypothetical protein			
	M422DRAFT_265232	7.03	1.03	1.12
g3485	hypothetical protein BC1G_05584	5.03	6.77	5.05
g3538	trypsin-like serine protease	6.32	3.37	5.52
g3540	hypothetical protein			
	GYMLUDRAFT_706765	4.94	5.41	7.85
g3834	kinase domain-containing	4.99	5.49	6.74
g3893	hypothetical protein GLAREA_08855	2.97	6.54	9.09
g3895	---NA---	3.04	3.78	5.60
g3985	general substrate transporter	5.42	6.96	5.64
g4056	---NA---	2.27	2.61	-2.19
g4166	hypothetical protein			
	AOQ84DRAFT_321579	5.55	5.69	2.96
	[Laccaria bicolor S238N-H82]			
g4217	XP_001878005	4.48	8.62	6.53
g4239	aat1p	3.34	4.31	-1.52
g4287	camp-mediated signaling	6.21	6.53	5.31
g4386	protein of unknown function DUF1996	6.16	5.76	4.09
g4444	hypothetical protein A103_09169	6.71	-1.91	-2.93
g4454	hypothetical protein	0.09	1.00	-1.89
g4480	hypothetical protein SS1G_00468	7.56	2.27	2.53
g4508	Scytalone dehydratase	5.79	6.86	8.66
g4509	1,3,6,8-tetra-HN reductase	3.70	5.71	9.10
g4510	transcription factor	5.18	5.95	7.89
g4511	non-reducing type I polyketide synthase	5.71	6.48	8.63
g4590	hypothetical protein PEX1_078220	6.62	6.63	5.05
g4604	---NA---	6.77	2.62	0.98
g4607	hypothetical protein			
	OIDMADRAFT_35369	0.37	4.33	8.17
g4652	toxin subunit	2.81	3.09	-1.82
g4653	toxin subunit , partial	3.33	3.91	0.86
g4702	Cytochrome P450	4.36	4.45	1.81

g4703	hypothetical protein FOVG_18288	4.67	4.50	1.97
g4718	---NA---	6.24	0.28	0.97
g4752	putative d-mandelate protein	6.29	6.43	7.79
g4875	---NA---	5.87	5.08	6.17
g4884	hypothetical protein ANI_1_1854104	2.11	6.25	7.14
g4922	hypothetical protein FVEG_12244	3.04	3.84	0.81
g4953	zinc knuckle transcription factor	6.70	2.99	2.90
g4995	hypothetical protein L228DRAFT_240605	3.70	4.12	0.29
g5073	NADP-dependent alcohol dehydrogenase 6	5.24	7.36	7.69
g5120	---NA---	6.36	6.65	5.61
g5145	nitroreductase	4.14	5.80	7.00
g5155	Uncharacterized transporter	3.92	5.86	2.41
g5156	hypothetical protein Z518_02103	4.14	6.03	2.55
g5157	MOSC domain containing	6.86	5.91	3.46
g5168	Subtilisin-like protease	6.42	1.60	0.00
g5248	hypothetical protein PTT_05645	6.54	-1.40	-1.43
g5388	transcription initiation factor iif	5.38	5.95	4.37
g5389	---NA---	4.42	5.88	4.16
g5401	Translation initiation factor IF-2	-3.23	0.32	-3.35
g5410	hypothetical protein VE04_07924	5.67	5.78	7.24
g5423	hypothetical protein MAJ_10788, partial	6.99	3.93	3.80
g5452	---NA---	7.75	2.35	2.50
g5476	hypothetical protein AOQ84DRAFT_431293	4.98	5.07	2.26
g5477	hypothetical protein AGABI2DRAFT_123004	4.58	4.23	1.57
g5590	subtilisin-like protein	5.32	4.97	2.75
g5744	---NA---	7.67	-0.16	0.91
g5832	carbonic anhydrase	7.12	7.40	5.48
g5846	---NA---	1.97	-3.72	-1.13
g5896	hypothetical protein CLCR_09305	4.11	4.93	6.70
g5967	hypothetical protein AZE42_02734	3.31	4.91	2.10
g6016	integral membrane	5.06	6.14	4.73
g6018	major facilitator superfamily transporter	5.53	6.92	4.52

g6019	DUF341 domain	3.42	5.99	2.61
g6021	reducing type I polyketide synthase	3.62	4.07	0.50
g6022	reducing type I polyketide synthase	4.05	4.52	1.63
g6026	hypothetical protein FE78DRAFT_72758	7.34	1.77	2.03
g6077	hypothetical protein MGG_02083	7.25	1.57	0.47
g6187	Oxoglutarate iron-dependent oxygenase degradation domain hypothetical protein	3.71	0.39	-2.99
g6220	RHOBADRAFT_67013	4.83	7.32	4.31
g6221	hypothetical protein ARAM_006588	3.93	5.51	2.18
g6278	Sugar transporter STL1	4.83	6.30	4.70
g6296	---NA--- hypothetical protein	2.85	6.33	3.30
g6304	CC77DRAFT_987110	6.52	5.15	3.37
g6386	hypothetical protein V490_07430	5.97	5.62	2.50
g6387	---NA--- S-adenosyl-L-methionine-dependent	1.65	-0.62	-3.87
g6501	methyltransferase	2.89	6.24	2.80
g6504	---NA---	3.93	-4.86	-2.45
g6505	---NA--- hypothetical protein	2.55	5.07	0.98
g6590	HMPREF1541_09740 [Laccaria bicolor S238N-H82]	6.19	8.44	8.32
g6633	XP_001878005	2.05	3.82	-3.55
g6697	hypothetical protein HIM_12619	-1.03	-0.68	-3.76
g6711	carbohydrate-binding module family 13	-1.01	7.67	4.74
g6747	barwin-like endoglucanase	4.17	6.05	3.70
g6750	---NA---	-1.16	4.37	1.11
g6756	---NA---	0.84	4.52	0.86
g6929	---NA---	7.10	2.45	2.35
g7123	MFS general substrate transporter	8.32	4.82	5.99
g7295	hypothetical protein XA68_3274	5.19	5.67	4.76
g7331	LAFE_OE15038g1_1	6.46	6.54	7.99
g7358	hypothetical protein A109_02403	3.74	7.08	5.00
g7359	Polysaccharide synthase Cps1	5.59	7.15	4.97
g7415	hypothetical protein MGYG_04240	6.24	6.04	7.64
g7418	---NA---	6.62	-2.11	-3.05

g7519	---NA---	-0.42	0.75	-1.91
	hypothetical protein			
g7528	ASPWEDRAFT_144002	1.77	4.96	1.49
	hypothetical protein			
g7530	LY89DRAFT_743438	2.80	3.42	0.43
g7641	complexed with Cdc5 Cwf16	7.80	6.00	4.17
	hypothetical protein			
g7654	L228DRAFT_267680	5.92	6.13	4.00
g7681	hypothetical protein H109_01118	1.75	2.13	5.19
	hypothetical protein			
g7747	BOTBODRAFT_118501	7.23	1.86	1.65
	hypothetical protein			
g7839	K441DRAFT_742369	2.43	3.04	-1.68
g7881	hypothetical protein PV08_05142	5.78	8.40	7.57
g8033	21 kDa subunit of NADH dehydrogenase	0.79	1.34	-3.08
g8039	c6 zinc finger domain containing	5.63	6.47	4.88
g8072	disaggregation chaperone	-5.06	-1.75	-5.67
	hypothetical protein			
g8075	BAUCODRAFT_156653	6.11	6.06	7.17
g8089	---NA---	5.58	6.48	4.84
g8150	hypothetical protein ACO22_01134	6.43	6.95	6.07
g8179	Collagen triple helix repeat-containing	1.67	-0.13	3.85
g8190	SGNH hydrolase	-0.44	-2.29	2.44
g8218	outer membrane autotransporter barrel	8.56	8.67	6.50
	hypothetical protein			
g8246	COCSADRAFT_140624	5.44	4.71	6.25
g8404	---NA---	0.37	1.26	-3.36
g8535	hydroxyindole o	6.50	7.84	6.15
g8548	hypothetical protein PV09_00215	5.00	6.28	3.93
g8565	WD40 repeat	1.00	-1.82	0.40
g8579	hypothetical protein V496_03516	6.35	5.92	3.58
g8700	small oligopeptide OPT family	6.02	6.90	3.98
g8765	hypothetical protein	6.04	4.08	6.99
g8781	multicopper oxidase	0.77	3.33	5.50
g8782	---NA---	1.79	3.92	-0.20
g8789	---NA---	-2.03	-1.59	-4.30
g8882	hypothetical protein	2.92	5.55	2.21

	OIDMADRAFT_105967			
g8903	hypothetical protein ASPCAL09681	1.64	3.41	-0.27
g8922	Nonribosomal Peptide Synthase (NRPS)	-2.64	0.07	-3.23
	hypothetical protein			
g8994	SCHPADRAFT_996133	6.26	5.67	2.18
	E3 ubiquitin ligase complex SCF subunit			
g8995	scon	6.57	6.14	1.29
	hypothetical protein			
g9075	L228DRAFT_263014	4.90	5.53	3.43
g9085	cytochrome P450	2.66	3.27	-1.43
g9086	terpene cyclase	2.78	1.66	-2.88
g9099	---NA---	1.48	1.21	-3.24
g9151	hypothetical protein H101_03408	7.15	2.57	2.58
	hypothetical protein			
g9159	HYPBUDRAFT_152817	3.67	5.23	2.76
g9160	RNA-binding domain-containing	4.15	5.64	2.98
	hypothetical protein			
g9162	ASPZODRAFT_126497	1.72	1.10	-2.71
g9168	na+ h+ exchanger 1	0.08	0.00	-3.57
g9171	sterol desaturase family	6.73	-0.35	-0.64
g9289	hypothetical protein TDEL_OA05340	3.78	5.36	3.15
	signal recognition particle, SRP19			
g9309	subunit	5.41	7.18	8.11
	Podospira anserina S mat genomic DNA			
g9319	chromosome supercontig 2	6.29	6.22	3.13
g9320	hypothetical protein XA68_103	6.22	6.16	2.85
g9346	Pre-rRNA-processing PNO1	2.39	5.43	2.76
g9390	---NA---	7.06	-0.33	-1.85
g9449	---NA---	7.87	2.53	1.89
g9517	scp-like extracellular	7.76	3.45	3.80
g9541	hypothetical protein EPUS_08791	3.64	-2.77	0.02
g9546	hypothetical protein PGUG_04206	-0.85	0.86	-2.04
g9562	polyketide synthase 34	6.00	4.89	7.20
	hypothetical protein			
g9736	ASPWEDRAFT_501621	1.70	4.26	1.14
g9751	mate family transporter	4.99	7.27	7.38
g9822	ATP-dependent RNA helicase ded1,	7.60	-3.64	-1.61

	hypothetical protein			
g9834	M408DRAFT_23609	5.66	7.00	5.64
g9882	amine flavin-containing superfamily	6.43	6.02	7.35
	hypothetical protein			
g10103	K441DRAFT_603565	5.93	5.60	6.99
g10144	hypothetical protein W97_09195	2.67	-2.52	0.83
g10178	---NA---	0.66	3.68	-1.17
g10252	polyketide synthase	0.08	0.73	-2.62
g10277	retinol dehydrogenase 12	5.57	8.52	8.30
	hypothetical protein			
g10296	AOQ84DRAFT_351571	4.20	4.03	7.33
	hypothetical protein			
g10304	CC84DRAFT_1102722	4.65	5.02	1.36
g10306	reducing type I polyketide synthase 10	1.39	0.40	-2.32
g10320	hypothetical protein AYL99_01801	3.17	3.74	0.84
g10516	secretory lipase	4.82	6.32	4.94
g10546	---NA---	3.19	3.27	5.41
	hypothetical protein			
g10555	EURHEDRAFT_526790	4.52	4.21	5.81
g10570	UPF0255 protein	3.28	4.74	6.17
g10688	carboxypeptidase S1	4.68	5.97	3.94
g10725	UPF0061-domain-containing protein	-1.38	2.06	-3.07
g10792	---NA---	8.24	6.14	5.76
g10793	p53-like transcription factor	8.07	5.77	5.52
g10851	---NA---	5.54	9.86	8.25
g10880	---NA---	7.49	-2.31	-2.67
g10919	Uncharacterized lipo ybbD P40406	3.25	3.78	0.29
g10948	---NA---	-0.80	0.77	-4.45
	hypothetical protein			
g10966	CALVIDRAFT_540720	6.61	2.15	2.13
g10970	hypothetical protein UCREL1_2371	1.79	6.24	7.42
g11004	galectin-like protein [Laccaria bicolor S238N-H82]	3.31	8.43	6.36
g11044	XP_001878005	6.47	6.06	7.16
g11053	predicted protein	4.95	7.17	8.16
g11054	uncharacterized protein A8A55_2761	4.24	6.83	7.65
g11057	Ubiquitin fusion degradation 4	2.45	5.15	6.56

g11138	Common central domain of tyrosinase	5.15	5.72	8.26
g11144	kinase-like domain	3.27	0.79	-2.16
g11206	molecular chaperone	-2.84	-0.50	-3.54
	hypothetical protein			
g11247	K441DRAFT_659689	4.20	4.70	6.44
g11259	MFS multidrug resistance transporter,	4.44	5.36	7.13
g11327	beta-lactamase transpeptidase	5.69	-5.73	-5.70
g11331	O-methyltransferase	3.07	7.64	7.26
g11372	predicted protein	-1.84	-1.33	-3.62
g11415	hypothetical protein FE78DRAFT_72758	6.85	1.16	1.10
g11420	carbohydrate-binding module family 50	3.11	-0.24	-3.19
g11434	uncharacterized protein PAC_17075	0.46	-2.20	5.70
g11435	uncharacterized protein PAC_10069	3.17	0.70	6.42
g11436	related to multidrug resistant	-1.22	0.68	5.35
g11437	uncharacterized protein PAC_17077	3.20	2.17	6.57
g11482	Metal regulatory transcription factor 1	1.96	4.07	-0.54
g11483	cytochrome p450	3.29	5.11	0.89
g11485	hypothetical protein BofuT4_P025320.1	4.77	5.97	3.32
g11486	cytochrome p450	3.52	5.15	1.60
g11615	heat shock 30	7.49	4.59	4.22
g11622	hypothetical protein V498_10729	3.47	3.28	5.57
	phospho-2-dehydro-3-deoxyheptonate			
g11692	aldolase, tyrosine-inhibited	6.31	8.02	8.39
g11695	heterokaryon incompatibility	2.85	1.60	4.95
g11711	HC-toxin efflux carrier TOXA	5.59	4.51	7.47
g11712	methyltransferase	5.48	2.00	7.75
	hypothetical protein			
g11713	M437DRAFT_59426	6.34	2.40	7.55
g11797	NAD(P)-binding protein	4.90	5.03	2.98
g11900	histidine kinase	1.85	3.55	-0.30
	MFS transporter, ACS family, allantate			
g11954	permease	3.39	6.12	2.86
g11995	carboxypeptidase cpdS	5.17	4.65	5.67
	hypothetical protein			
g12020	LY89DRAFT_747417	6.17	5.71	3.09
g12045	serine threonine kinase	5.48	4.72	1.47
g12046	hypothetical protein	3.71	3.35	-1.62

BATDEDRAFT_24227				
g12048	hypothetical protein XA68_2246	5.27	6.27	2.80
g12049	fructose-bisphosphate aldolase,	5.01	5.81	2.80
g12050	---NA---	5.54	6.59	3.61
g12052	---NA---	5.34	6.44	3.69
g12077	---NA---	-1.60	4.10	-1.04
g12122	---NA---	1.72	1.24	4.51
g12125	---NA---	-0.35	5.40	6.98
g12126	---NA---	1.62	4.49	6.32
g12222	FAS1 domain-containing	3.79	7.80	8.86
g12259	predicted protein	0.35	-0.04	-2.34
g12283	AP endonuclease	3.59	5.28	6.52
g12284	hypothetical protein PV04_01423	4.33	6.14	7.31
	hypothetical protein			
g12306	M404DRAFT_122096	3.43	3.00	-1.79
g12402	ankyrin, partial	0.56	0.25	-1.77
g12409	hypothetical protein ASPCAL07495	2.56	6.23	3.92
g12420	hypothetical protein CGLO_08849	4.18	0.31	-2.52
g12432	---NA---	5.71	4.47	-0.31
g12460	---NA---	-0.23	4.30	0.11
g12505	solid-state culture specific	5.55	3.76	6.26
g12554	---NA---	4.19	6.04	4.34
g12567	hypothetical protein PV07_04705	5.70	6.03	7.47
g12615	---NA---	7.98	10.73	10.64
g12771	amino acid transporter	5.80	6.75	4.81
g12863	---NA---	5.11	0.51	-1.97
g12965	hypothetical protein FE78DRAFT_72758	7.62	3.14	3.26
g12991	uncharacterized protein FFUJ_00300	2.57	1.35	5.27
g12994	predicted protein	2.84	8.13	9.40
g12999	integral membrane	0.72	3.23	-0.68
g13045	---NA---	3.46	4.56	1.84
g13046	alpha beta-Hydrolase	6.32	7.20	5.47
g13127	---NA---	2.76	3.36	-0.42
g13147	related to ferroportin 1	3.47	3.69	0.06
g13150	hypothetical protein ASPCAL07495	1.65	2.99	-3.21
g13153	hypothetical protein	4.87	6.36	5.54

	LY89DRAFT_783720			
g13209	alpha beta-hydrolase	4.44	7.43	7.43
g13255	---NA---	2.88	5.89	3.71
	hypothetical protein			
g13256	ASPGLDRAFT_182435	5.07	3.73	7.79
g13257	catalytic	5.08	4.36	7.71
g13258	C6 zinc finger domain	3.62	2.55	5.03
g13275	---NA---	0.46	0.67	-1.98
g13314	---NA---	0.36	0.48	-2.03
g13317	hypothetical protein EPUS_06272	7.29	3.39	2.12
g13326	Gag-Pol poly	-0.48	0.35	-2.60
g13376	hypothetical protein sscl_09g068980	3.47	4.29	0.01
g13377	MFS transporter	3.87	4.24	0.01
g13413	hypothetical protein AYO22_09380	5.11	6.40	4.72
g13423	---NA---	6.92	5.21	6.82
g13424	thermosensitive gluconokinase	6.80	5.40	6.91
g13440	carbohydrate-binding module family 32	2.38	3.01	5.34
	Mediator of RNA polymerase II			
g13475	transcription subunit 13	5.59	6.48	4.50
g13486	hypothetical protein PV07_12346	2.31	3.22	7.38
g13487	beta-ig-h3 fasciclin	1.84	5.12	8.67
g13492	---NA---	6.29	0.73	1.00
g13505	---NA---	3.36	-5.40	-6.36
g13508	cell division control 42	6.21	-4.90	-1.41
g13512	hypothetical protein AYO20_06870	5.14	7.01	5.52
g13515	LNS2-domain-containing protein	1.65	1.30	-2.19
	hypothetical protein			
g13593	LY89DRAFT_609882	2.73	3.91	0.08
g13713	hypothetical protein EPUS_05087	6.86	4.79	6.72
g13715	glyoxylate pathway regulator	4.54	6.76	5.55
	hypothetical protein			
g13723	SISNIDRAFT_484147	-0.33	5.45	6.90
	hypothetical protein			
g13734	K432DRAFT_290587	3.23	2.60	6.63
g13749	FAD dependent oxidoreductase	5.19	7.88	7.75
g13757	TPR domain-containing	3.72	-4.97	-4.95
g13783	---NA---	1.02	2.07	-2.28

g13786	gaba permease	4.16	3.97	5.96
g13788	hypothetical protein VF21_00684	3.36	0.31	5.39
	hypothetical protein			
g13789	AOQ84DRAFT_329710	3.88	1.56	5.61
g13820	hypothetical protein W97_07655	6.92	1.67	1.23
	hypothetical protein			
g13837	M413DRAFT_32936	8.05	1.57	0.76
	High-affinity methionine permease			
g13841	P50276	1.75	3.41	6.50
g13918	---NA---	4.91	7.16	5.76
g13967	---NA---	7.22	4.42	1.81
g13968	RNA-binding protein	7.37	4.95	2.94
g14007	hypothetical protein AZE42_07563	7.44	6.34	4.16
g14069	hypothetical protein V490_04153	0.58	0.63	-1.85
g14086	---NA---	0.89	3.79	-0.36
g14096	peptidylprolyl isomerase	4.11	6.06	6.87
	hypothetical protein			
g14126	K441DRAFT_661295	-0.93	-2.34	1.70
g14129	alpha beta-hydrolase	3.92	-3.02	-4.88
g14215	hypothetical protein HIM_07753	-0.01	-2.87	0.74
	hypothetical protein			
g14225	OGAPODRAFT_12856	6.85	1.80	1.82
g14248	---NA---	5.85	6.38	1.81
	hypothetical protein			
g14275	AOQ84DRAFT_380795	3.23	3.39	6.77
g14364	PenC	-1.43	0.34	4.55
	S-adenosyl-L-methionine-dependent			
g14365	methyltransferase	-1.17	-1.49	4.31
	hypothetical protein			
g14452	HYPsudRAFT_36721	6.50	-2.27	-2.35
g14467	hypothetical protein MBM_00672	6.16	5.22	1.94
g14470	hypothetical protein A4X13_g7656	0.51	1.39	-2.21
g14474	---NA---	4.88	8.03	8.79
	hypothetical protein			
g14555	M422DRAFT_39030	3.64	6.71	6.69
g14560	hypothetical protein PGRI_095850	6.61	2.41	3.12
g14603	predicted protein	0.02	5.18	0.11
g14629	hypothetical protein W97_08269	5.75	7.81	8.21

	hypothetical protein			
g14674	ANO14919_024120	5.33	6.19	4.48
	hypothetical protein			
g14676	MYCFIDRAFT_212921	4.79	6.36	5.39
g14717	---NA---	6.83	2.77	2.24
g14736	hypothetical protein Hypma_11659	1.87	5.70	6.99
	hypothetical protein			
g14762	K441DRAFT_650523	4.81	7.85	5.97
g14780	---NA---	4.59	4.55	1.66
	hypothetical protein			
g14802	AOQ84DRAFT_191815	-0.11	6.31	4.49
g14803	---NA---	2.90	5.82	1.62
g14849	hypothetical protein V496_02147	1.63	0.74	-2.53
	hypothetical protein A1Q2_08520			
g14931	(mitochondrion)	8.40	9.12	7.18
g14932	hydrolase, partial	8.33	9.49	7.23
	hypothetical protein			
g14933	CONCODRAFT_139759	8.28	9.52	7.19
g14956	---NA---	2.89	4.87	0.24
g14998	---NA---	7.49	-0.84	-0.24
g15026	ZIP zinc iron transport family	6.12	4.62	6.39
g15072	hypothetical protein PTTG_27154	-1.70	4.56	8.29
g15092	hypothetical protein PFICI_07720	3.37	4.68	1.88
g15231	---NA---	5.49	7.82	8.39
g15278	---NA---	7.01	1.98	3.82
g15390	---NA---	4.30	6.93	7.05
g15421	sporulation associated	2.47	4.63	2.33
g15430	---NA---	7.34	0.23	-0.74
	hypothetical protein			
g15478	HMPREF1541_07223	5.67	6.50	4.80
g15501	hypothetical protein PFICI_15058	6.61	4.38	-0.67
g15502	lectin-2	7.95	6.12	1.75
g15614	GroES-like protein	3.78	8.05	7.26
g15640	hypothetical protein PTT_10076	7.85	8.56	6.56
g15641	---NA---	8.37	9.45	7.48
g15651	hydrolase, partial	6.62	2.56	3.85
g15670	hydrolase, partial	7.26	3.69	3.89

g15702	hypothetical protein I308_02452	1.26	0.38	-3.76
g15731	Lysine-specific demethylase 5A	5.72	6.12	4.80
g15759	integral membrane	3.14	4.47	1.38
g15914	---NA---	5.90	0.96	3.43
	hypothetical protein			
g15938	ASPGLDRAFT_22819	6.56	-0.39	0.37
g15995	MFS transporter	3.43	6.05	2.79
g16044	polyketide synthase	4.87	5.72	1.00
g16045	hypothetical protein DHEL01_09803	5.55	6.38	3.10
g16121	hypothetical protein, partial	6.68	3.12	5.80
g16178	glutathione S-transferase	7.26	5.34	3.53
g16199	---NA---	5.70	5.25	3.89
g16231	---NA---	8.02	9.65	7.49
g16255	Sporulation-specific 5	6.93	3.34	1.43
g16262	hypothetical protein CORT_OF02850	1.65	3.83	-1.50
g16282	mfs multidrug transporter	6.61	7.08	4.56
g16287	---NA---	5.07	6.64	8.09
g16309	permease, partial	7.11	2.93	4.48
	hypothetical protein			
g16379	ANO14919_071370	2.32	3.50	-0.20

Table S20: List of 265 DEGs between A (au-09) vs. TW (au-09), and the 56-background variation DEGs are highlighted in red.

Gene id.	Description	A (au-2009) (log2 RPM)	TW (au-2009) (log2 RPM)
g135	hypothetical protein SPPG_07180	6.69	2.97
g199	hypothetical protein ZTR_09325	2.38	6.64
g239	Dehydroquinase synthase	6.41	9.33
g246	---NA---	6.85	0.56
g336	---NA---	9.56	6.99
g340	FACT complex subunit spt16	7.20	3.38
g581	ABC transporter	9.08	6.38
g689	---NA---	6.76	0.60
g766	hypothetical protein W97_07533	6.72	3.47
g816	hypothetical protein TGAM01_04190	7.33	3.46
g1006	hypothetical protein VF21_09233	5.32	8.12
g1014	hypothetical protein W97_02484	5.72	-1.10
g1044	vanillyl alcohol oxidase	4.66	7.91
g1045	amp-binding enzyme	7.78	11.49
g1049	hypothetical protein COCC4DRAFT_37502	4.10	7.82
g1050	2og-fe oxygenase family oxidoreductase	2.93	8.13
g1051	hypothetical protein HIM_05906	3.55	9.10
g1119	hypothetical protein W97_07378	6.40	-2.17
g1184	MFS general substrate transporter	8.76	5.73
g1256	FAD dependent oxidoreductase superfamily	8.18	5.08
g1322	hypothetical protein ANI_1_2290184	2.18	8.52
g1424	glycoside hydrolase family 32	6.37	8.42
g1522	hypothetical protein W97_08454	7.32	1.97
g1523	M repeat	4.56	-1.67
g1639	Piwi-domain-containing protein	8.43	2.92
g1670	hypothetical protein IQ07DRAFT_590219	3.42	6.56
g1684	---NA---	9.66	6.03
g1793	---NA---	4.36	-2.17
g2053	---NA---	10.32	5.85

g2080	---NA---	6.48	-2.25
g2113	hypothetical protein W97_03127	9.68	3.75
g2226	hypothetical protein NFIA_022830	7.53	3.90
g2246	hypothetical protein AJ78_01803	10.10	7.13
g2371	hypothetical protein AOCH_007280, partial	9.71	6.85
g2651	related to dis1-suppressing kinase dsk1	5.13	-0.14
g2672	hypothetical protein K432DRAFT_400307	6.08	2.46
g2718	hypothetical protein	6.77	0.86
g2749	hypothetical protein BAUCODRAFT_27901	7.17	1.15
g2750	hypothetical protein V493_01779	6.81	1.60
g2870	hypothetical protein TI39_contig427g00003	6.27	1.90
g2963	hypothetical protein MVLG_00710	5.41	1.82
g3121	---NA---	9.29	7.15
g3282	hypothetical protein FOYG_00325	6.66	2.33
g3496	Fungal transcriptional regulatory	6.65	3.26
g3576	hypothetical protein K432DRAFT_383967	7.19	9.43
g3594	---NA---	4.64	-1.67
g3702	Tyrosinase Q92396	6.06	8.19
g4105	glycoside hydrolase family 16	5.47	0.92
g4217	[Laccaria bicolor S238N-H82] XP_001878005	3.98	7.70
g4440	putative dynamin	5.60	-0.44
g4444	hypothetical protein A103_09169	6.88	0.19
g4505	carbohydrate esterase family 16	6.09	2.61
g4558	haus augmin-like complex subunit 8 isoform x1	7.60	4.91
g4586	hypothetical protein EURHEDRAFT_376324	7.94	4.10
g4604	---NA---	7.34	4.19
g4718	---NA---	8.28	5.86
g4801	---NA---	5.95	-0.67
g4822	aquaporin-like protein	6.75	3.47
g4884	hypothetical protein ANI_1_1854104	4.97	8.55
g4953	zinc knuckle transcription factor	8.40	4.05
g4977	---NA---	6.65	1.95
g4983	hypothetical protein CSIM01_12475	3.75	-2.25
g5001	hypothetical protein T310_1084	7.05	3.94
g5168	Subtilisin-like protease	6.56	1.39
g5248	hypothetical protein PTT_05645	5.20	-2.25
g5284	DNA mismatch repair , core	8.00	4.96

g5367	hypothetical protein EPUS_08071	7.61	3.73
g5401	Translation initiation factor IF-2	-1.07	6.26
g5423	hypothetical protein MAJ_10788, partial	6.55	3.18
g5452	---NA---	9.11	4.94
g5466	FAD FMN-containing isoamyl alcohol oxidase	6.97	9.04
g5538	ubiquitin homeostasis lub1	6.49	3.56
g5562	hypothetical protein SPPG_00479	2.68	5.81
g5719	UDPglucose 6-dehydrogenase	6.76	1.15
g5733	mitochondrial precursor	6.59	8.29
g5744	---NA---	7.06	-0.14
g5801	hypothetical protein ALT_4500	5.60	0.49
g5836	predicted protein	4.62	-2.17
g5846	---NA---	3.94	-2.25
g5867	m repeat	8.23	5.66
g5912	hypothetical protein AOQ84DRAFT_124757	7.70	5.11
g5976	---NA---	7.01	3.51
g5984	hypothetical protein W97_01113	5.77	0.89
g6048	related to alpha-1,2-mannosyltransferase	9.31	7.23
g6077	hypothetical protein MGG_02083	7.55	-0.10
g6137	hypothetical protein COCMIDRAFT_34414	5.16	-1.17
g6151	glycoside hydrolase family 17	8.99	6.33
g6246	Tetratricopeptide-like helical	5.48	0.78
g6269	hypothetical protein TEQG_07362	9.71	5.58
g6445	hypothetical protein H072_1194	5.24	8.44
g6486	P-loop containing nucleoside triphosphate hydrolase S-adenosyl-L-methionine-dependent methyltransferase	10.09	7.85
g6501	---	4.64	7.13
g6504	---NA---	6.00	1.06
g6527	uncharacterized protein RSE6_05777	6.82	2.45
g6717	hypothetical protein GLAREA_11641	5.77	-1.67
g6718	hypothetical protein I206_06573	5.73	0.99
g6784	Transcription factor TFIIIB related	7.06	1.04
g6836	Aspartic-type endopeptidase ()	6.61	2.29
g6853	hypothetical protein FG05_01754	4.07	7.53
g6855	ribonuclease H	9.10	6.82
g6929	---NA---	10.08	8.21
g6978	---NA---	4.82	0.63

g7085	TPA_exp: Tyrosinase	8.30	4.59
g7123	MFS general substrate transporter	10.12	7.92
g7223	predicted protein	9.70	7.71
g7298	---NA---	5.76	1.76
g7299	conserved hypothetical protein	4.81	-1.17
g7301	GMC oxidoreductase	7.26	3.80
g7359	Polysaccharide synthase Cps1	6.06	8.32
g7399	hypothetical protein W97_01699	4.34	-1.67
g7418	---NA---	6.91	-2.25
g7501	ribonuclease h	6.05	0.81
g7545	kinesin light chain	7.50	9.52
g7566	vacuolar carboxypeptidase Cps1	7.25	3.97
g7632	mitochondrial glyco	4.57	-1.67
g7807	---NA---	6.62	1.57
	2,3-bisphosphoglycerate-independent		
g8066	phosphoglycerate mutase	-2.15	4.51
g8070	---NA---	-2.07	3.92
g8071	dinitrogenase reductase, partial	-2.07	4.21
g8072	disaggregation chaperone	-2.15	4.56
g8089	---NA---	6.48	8.38
g8164	cytochrome P450	4.67	7.54
g8166	reducing type I polyketide synthase	2.11	5.85
g8167	reducing type I polyketide synthase	1.41	5.73
g8168	reducing type I polyketide synthase	4.54	8.42
g8169	reducing type I polyketide synthase	2.52	6.84
g8217	RING finger	4.75	-1.10
g8261	low-affinity ammonium transporter	5.92	7.91
g8396	hypothetical protein CONLIGDRAFT_432127	5.42	1.79
g8535	hydroxyindole o	7.41	9.32
g8787	allophycocyanin subunit beta	-0.33	5.20
g8788	---NA---	0.20	6.10
g8789	---NA---	-0.49	5.98
g8882	hypothetical protein OIADMADRAFT_105967	4.64	7.15
g8897	---NA---	8.85	6.39
g9019	hypothetical protein L228DRAFT_284146	6.86	3.26
g9041	related to major facilitator (MFS1) transporter	8.14	6.05
g9153	---NA---	7.08	3.18

g9170	hypothetical protein	5.55	-1.17
g9196	hypothetical protein V496_02605	3.19	8.48
g9197	chromatin assembly factor 1 subunit b	-2.15	3.87
g9270	---NA---	5.57	-0.37
g9279	nuclear and cytoplasmic polyadenylated RNA-binding , partial	4.22	-0.94
g9300	---NA---	5.34	-2.25
g9318	hypothetical protein RO3G_02968	4.40	-2.25
g9366	---NA---	7.60	3.07
g9390	---NA---	6.14	-0.60
g9449	---NA---	8.78	4.10
g9517	scp-like extracellular	8.06	3.54
g9532	hypothetical protein GGTG_03837	7.07	3.97
g9556	hypothetical protein AOQ84DRAFT_421258	4.24	7.20
g9578	hypothetical protein COCSADRAFT_100576	7.67	4.52
g9822	ATP-dependent RNA helicase ded1,	9.56	0.02
g9913	hypothetical protein L228DRAFT_170648	5.57	8.22
g9934	hypothetical protein L228DRAFT_270124	7.54	5.03
g9947	ARM repeat-containing hypothetical protein AOQ84DRAFT_271286,	8.02	4.17
g10033	partial	7.69	4.46
g10212	cytochrome P450	1.25	6.99
g10213	cytochrome P450	0.17	5.41
g10214	short chain dehydrogenase reductase	-1.20	5.39
g10367	hypothetical protein MPH_10379	4.55	7.20
g10419	hypothetical protein AUEXF2481DRAFT_88666	7.81	9.70
g10614	hypothetical protein MGYG_06716 26S proteasome non-ATPase regulatory subunit	4.75	-1.67
g10746	10	6.51	2.51
g10880	---NA---	7.40	-2.25
g10970	hypothetical protein UCREL1_2371	1.09	5.68
g11003	---NA---	4.55	-2.25
g11004	galectin-like protein	5.04	10.02
g11086	Ferritin-like protein	5.62	8.41
g11170	related to lipoxxygenase 1	-0.83	5.37
g11171	Lipoxxygenase	0.57	6.19
g11184	transitional endoplasmic reticulum ATPase	0.30	7.46
g11234	hypothetical protein PV05_07258	5.18	1.11

g11262	---NA---	7.42	-2.25
g11295	phycocyanin subunit beta	-1.20	6.18
g11296	phycocyanin subunit alpha	-1.57	5.43
g11303	ubiquitin carboxyl-terminal hydrolase 2	6.83	4.04
g11327	beta-lactamase transpeptidase	8.63	1.69
g11331	O-methyltransferase	2.65	6.70
g11615	heat shock 30	7.32	3.49
g11749	hypothetical protein TSTA_020850	7.78	2.74
g11867	ATP synthase subunit alpha, mitochondrial	-2.07	5.16
g11868	ATP synthase F0 subunit B	-2.15	4.24
g11893	hypothetical protein MAC_08995	7.30	4.01
g12077	---NA---	-0.17	5.80
g12222	FAS1 domain-containing	4.47	8.72
g12261	hypothetical protein EPUS_07439	6.90	9.03
g12284	hypothetical protein PV04_01423	5.84	8.59
g12394	hypothetical protein W97_07005	7.19	3.36
g12432	---NA---	4.12	-1.67
g12474	ankyrin repeat-containing	5.21	0.19
g12494	exo-beta- -glucanase	-0.70	3.91
g12527	---NA---	6.65	2.66
g12629	---NA---	8.90	6.72
g12696	Cytochrome P450 pisatin demethylase	6.31	-0.88
g12697	integral membrane	4.34	-2.25
g12863	---NA---	4.47	-0.94
g12945	NAD(P)-binding Rossmann-fold containing	5.31	-0.88
g12994	predicted protein	1.91	7.12
g13001	hypothetical protein AOQ84DRAFT_337496, partial	5.20	1.42
g13060	laccase-like multicopper oxidase	9.40	11.47
g13150	hypothetical protein ASPCAL07495	2.09	6.24
g13196	RNI-like protein	5.37	-2.17
g13209	alpha beta-hydrolase	6.90	9.04
g13330	hypothetical protein EMPG_15460	7.51	4.46
g13337	Calcium calmodulin-dependent kinase type	1.31	6.76
g13487	beta-ig-h3 fasciclin	1.91	6.98
g13491	hypothetical protein SERLADRAFT_431820	4.49	-1.60
g13492	---NA---	6.28	1.83

g13494	---NA---	4.27	-1.67
g13506	mating-type MAT alpha 1	6.02	3.15
g13507	---NA---	4.17	-2.25
g13508	cell division control 42	8.32	-2.25
g13509	Pkinase-domain-containing protein	5.93	-2.25
g13639	protein TIF31	4.30	-1.10
g13695	uncharacterized protein PAC_01586	8.92	6.16
g13749	FAD dependent oxidoreductase	4.70	7.14
g13782	photosystem I core	2.56	7.63
g13783	---NA---	2.20	6.34
g13820	hypothetical protein W97_07655	6.27	-0.51
g13830	hypothetical protein V492_04989	5.83	8.10
g13837	hypothetical protein M413DRAFT_32936	7.50	2.83
g13841	High-affinity methionine permease P50276	4.60	8.27
g13856	MFS general substrate transporter	6.13	-2.25
g13857	MFS general substrate transporter	5.93	-1.38
g13957	hypothetical protein SMAC_02332	5.58	0.90
g14053	cytochrome p450	7.29	2.31
	RNA-directed DNA polymerase (reverse		
g14071	transcriptase)	3.94	-2.25
g14129	alpha beta-hydrolase	6.03	-2.25
g14173	hypothetical protein SS1G_01102	8.12	3.04
g14404	hypothetical protein VE00_05995	4.09	-2.25
g14424	High-affinity methionine permease P50276	3.92	7.59
g14452	hypothetical protein HYPsUDRAFT_36721	6.08	-2.25
g14519	integral membrane	1.46	6.24
g14561	---NA---	7.44	3.33
g14580	rna recognition domain-containing	7.16	2.15
g14596	Peptidase S8, subtilisin, Ser-active site	4.44	-2.25
g14717	---NA---	6.85	2.62
g14729	reducing type I polyketide synthase	3.97	-2.25
g14736	hypothetical protein Hypma_11659	0.90	6.55
g14802	hypothetical protein AOQ84DRAFT_191815	-0.99	5.23
g14910	hypothetical protein BAUCODRAFT_145551	5.89	8.07
	retrovirus-related pol poly from transposon tnt		
g14982	1-94	7.69	9.67
g14989	hypothetical protein PV09_01633	6.13	8.40
g15000	hypothetical protein ST47_g8211	6.73	-2.25

g15056	related to dis1-suppressing kinase dsk1	5.98	0.22
g15061	hypothetical protein AGABI2DRAFT_176739	7.54	9.83
g15122	WD40 repeat	4.93	-0.01
g15153	benzoate 4-monooxygenase cytochrome P450	6.21	8.47
g15177	hypothetical protein ALT_3415	4.26	-2.25
g15202	reducing type I polyketide synthase	6.57	-2.25
g15232	---NA---	4.69	7.24
	hypothetical protein ASPVEDRAFT_41603,		
g15272	partial	4.43	-2.25
g15430	---NA---	6.79	0.89
g15467	reducing type I polyketide synthase	4.67	-2.25
g15537	hypothetical protein EPUS_08934	3.74	-2.25
g15656	benzoate 4-monooxygenase cytochrome P450	5.53	8.04
g15819	uncharacterized protein PAC_19130	5.73	8.07
g15875	reducing type I polyketide synthase	4.14	-2.25
g15938	hypothetical protein ASPGLDRAFT_22819	7.24	-2.17
g16168	hypothetical protein L228DRAFT_242344	6.54	2.19
g16184	---NA---	4.12	-2.25
g16282	mfs multidrug transporter	6.96	-2.25
g16287	---NA---	3.94	7.32
g16358	hypothetical protein LY89DRAFT_739923	4.70	-2.25

Table S21: List of 219 DEGs between A (au-09) vs. R (au-10).

Gene id.	Description	A (au-2009) (log2 RPM)	R (au-2010) (log2 RPM)
g14	hypothetical protein GLAREA_03922	1.54	5.78
g199	hypothetical protein ZTR_09325	2.38	7.84
g246	---NA---	6.85	0.90
g336	---NA---	9.56	6.88
g340	FACT complex subunit spt16	7.20	3.58
g530	conidial pigment biosynthesis oxidase	4.77	8.19
g548	Abr1 brown 1	8.96	6.38
g581	Choline transport	9.08	6.06
g603	ABC transporter	6.82	4.21
g689	---NA---	6.76	0.26
g816	---NA---	7.33	3.88
g1014	hypothetical protein TGAM01_04190	5.72	1.61
g1052	hypothetical protein W97_02484	10.88	5.78
g1119	ABC transporter	6.40	-2.74
g1184	hypothetical protein W97_07378	8.76	6.24
g1318	MFS general substrate transporter	5.50	8.01
g1522	hypothetical protein PV05_01060	7.32	3.87
g1584	hypothetical protein W97_08454	6.39	8.34
g1639	uncharacterized protein PAC_09567	8.43	2.52
g1677	Piwi-domain-containing protein	9.86	8.22
g1684	dead deah box dna helicase	9.66	5.22
g1793	---NA---	4.36	-2.16
g1817	---NA---	6.46	2.01
g2053	vtc domain-containing	10.32	5.73
g2080	---NA---	6.48	-3.90
g2113	hypothetical protein W97_03127	9.68	3.97
g2226	hypothetical protein NFIA_022830	7.53	4.54
g2246	hypothetical protein AJ78_01803	10.10	6.48
g2371	hypothetical protein AOCH_007280,	9.71	7.24

	partial		
g2492	serine kinase	6.42	-2.74
g2494	---NA---	4.27	-2.74
g2497	conserved hypothetical protein	7.07	2.84
g2498	hypothetical protein EPUS_03937	7.75	5.43
g2718	hypothetical protein	6.77	1.58
g2749	hypothetical protein BAUCODRAFT_27901	7.17	-0.16
g2750	hypothetical protein V493_01779	6.81	2.04
g2784	hypothetical protein ASPZODRAFT_19280	5.02	0.58
	hypothetical protein		
g2870	TI39_contig427g00003	6.27	0.06
g2882	C6 transcription factor	8.41	4.69
g3172	hypothetical protein V500_00208	7.32	3.80
g3453	hypothetical protein M422DRAFT_265232	6.35	0.90
g3496	Fungal transcriptional regulatory	6.65	3.19
g3506	hypothetical protein EPUS_02122	8.64	5.60
g3576	hypothetical protein K432DRAFT_383967	7.19	9.32
g3594	---NA---	4.64	-3.74
g3702	Tyrosinase Q92396	6.06	8.09
g3893	hypothetical protein GLAREA_08855	1.30	5.43
g4105	glycoside hydrolase family 16	5.47	0.96
	hypothetical protein		
g4236	SCLCIDRAFT_1212282	5.40	1.06
g4440	putative dynamin	5.60	0.42
g4444	hypothetical protein A103_09169	6.88	-3.90
g4480	hypothetical protein SS1G_00468	7.09	1.90
g4505	carbohydrate esterase family 16	6.09	1.68
g4509	1,3,6,8-tetra-HN reductase	3.95	8.04
	haus augmin-like complex subunit 8		
g4558	isoform x1	7.60	4.91
	hypothetical protein		
g4586	EURHEDRAFT_376324	7.94	4.28
g4604	---NA---	7.34	2.76
g4607	hypothetical protein OI DMADRAFT_35369	-0.33	6.46
g4718	---NA---	8.28	2.28
g4801	---NA---	5.95	1.26

g4884	hypothetical protein ANI_1_1854104	4.97	8.78
g4953	zinc knuckle transcription factor	8.40	4.66
g4977	---NA---	6.65	1.78
g5001	hypothetical protein T310_1084	7.05	3.68
g5157	MOSC domain containing	8.58	6.42
g5248	hypothetical protein PTT_05645	5.20	-3.90
g5284	DNA mismatch repair , core	8.00	2.36
g5288	hypothetical protein MBM_01608	5.74	0.06
g5428	hypothetical protein EPUS_06503	5.63	0.84
g5452	---NA---	9.11	3.55
g5719	UDPglucose 6-dehydrogenase	6.76	2.85
g5744	---NA---	7.06	0.16
g5836	predicted protein	4.62	-1.74
g5912	hypothetical protein AOQ84DRAFT_124757	7.70	3.86
g6020	reducing type I polyketide synthase	8.18	5.59
g6026	hypothetical protein FE78DRAFT_72758	8.53	3.84
g6077	hypothetical protein MGG_02083	7.55	1.50
g6137	hypothetical protein COCMIDRAFT_34414	5.16	-1.16
g6269	hypothetical protein TEQG_07362	9.71	5.44
g6386	hypothetical protein V490_07430	6.67	1.38
g6403	nitrate reductase [NADPH]	5.64	7.95
g6445	hypothetical protein H072_1194	5.24	7.52
g6486	P-loop containing nucleoside triphosphate hydrolase	10.09	7.34
g6527	uncharacterized protein RSE6_05777	6.82	2.50
g6576	benzoate 4-monooxygenase cytochrome P450	7.74	5.11
g6717	hypothetical protein GLAREA_11641	5.77	-3.74
g6784	Transcription factor TFIIIB related	7.06	-0.94
g6867	hypothetical protein	5.00	-3.74
g6929	---NA---	10.08	5.45
g6947	P450 monooxygenase	1.67	6.30
g6948	benzoate 4-monooxygenase cytochrome	0.75	6.30
g7081	hypothetical protein V500_01722	4.60	-0.42
g7085	TPA_exp: Tyrosinase	8.30	2.34

g7123	MFS general substrate transporter	10.12	7.52
g7265	Collagen alpha-5(VI) chain	9.36	5.72
g7301	GMC oxidoreductase	7.26	4.55
g7377	NAD+ kinase	7.97	5.96
g7418	---NA---	6.91	-3.90
g7501	ribonuclease h	6.05	0.16
g7641	complexed with Cdc5 Cwf16	8.14	4.85
g7695	---NA---	5.86	-0.42
g7708	---NA---	6.06	2.01
g7752	hypothetical protein FISHEDRAFT_24821, partial	0.99	5.45
g7769	hypothetical protein UA08_03758	3.03	6.39
g7807	---NA---	6.62	2.26
g7952	hypothetical protein L228DRAFT_269213	7.10	3.74
g8182	WD40 repeat	7.31	4.80
g8218	outer membrane autotransporter barrel	8.06	3.91
g8838	SGNH hydrolase	3.94	7.65
g8897	---NA---	8.85	5.72
g9153	---NA---	7.08	2.59
g9155	Ribonuclease H1	4.46	-2.74
g9170	hypothetical protein	5.55	-0.04
g9196	hypothetical protein V496_02605	3.19	8.38
g9300	---NA---	5.34	-3.90
g9366	---NA---	7.60	4.20
g9390	---NA---	6.14	-2.16
g9449	---NA---	8.78	3.72
g9517	scp-like extracellular	8.06	2.75
g9548	hypothetical protein FE78DRAFT_72758	7.95	5.19
g9578	hypothetical protein COCSADRAFT_100576	7.67	3.32
g9700	hypothetical protein V491_01593	7.06	4.34
g9822	ATP-dependent RNA helicase ded1,	9.56	0.65
g9947	ARM repeat-containing	8.02	3.70
g10181	hypothetical protein AK830_g408	5.60	7.98
g10296	hypothetical protein AOQ84DRAFT_351571	4.16	8.46

g10332	hypothetical protein PV08_05721	5.67	7.58
g10421	reducing type I polyketide synthase	6.94	8.75
g10614	hypothetical protein MGYG_06716	4.75	-3.90
g10746	26S proteasome non-ATPase regulatory subunit 10	6.51	2.58
g10793	p53-like transcription factor	9.84	7.65
g10880	---NA---	7.40	-2.74
g10966	hypothetical protein CALVIDRAFT_540720	7.15	3.39
g10970	hypothetical protein UCREL1_2371	1.09	7.38
g11003	---NA---	4.55	-3.90
g11004	galectin-like protein	5.04	8.65
g11053	predicted protein	5.68	8.16
g11054	uncharacterized protein A8A55_2761	4.30	6.82
g11170	related to lipoxygenase 1	-0.83	5.61
g11171	Lipoxygenase	0.57	6.11
g11174	hypothetical protein K441DRAFT_670516	7.75	9.75
g11186	Ubiquitin-like protein	5.01	-3.90
g11262	---NA---	7.42	1.81
g11327	beta-lactamase transpeptidase	8.63	-3.74
g11415	hypothetical protein FE78DRAFT_72758	8.53	3.69
g11436	related to multidrug resistant	0.30	5.58
g11615	heat shock 30	7.32	4.35
g11622	hypothetical protein V498_10729	4.04	7.42
g11627	hypothetical protein EPUS_00425	-1.07	7.05
g11738	hypothetical protein SMAC_00573	2.04	6.62
g11749	hypothetical protein TSTA_020850	7.78	3.45
g11772	green fluorescent family	9.65	12.17
g11893	hypothetical protein MAC_08995	7.30	4.37
g12020	hypothetical protein LY89DRAFT_747417	8.52	5.13
g12021	uncharacterized protein PAC_08291	9.19	5.21
g12023	hypothetical protein CPC735_009070	0.46	6.15
g12046	hypothetical protein BATDEDRAFT_24227	4.55	-0.42
g12126	---NA---	2.39	6.38
g12169	flavin-binding monooxygenase	2.87	6.62
g12170	1, paxillin biosynthesis enzyme PAXC	4.60	8.08

g12222	FAS1 domain-containing	4.47	8.16
g12261	hypothetical protein EPUS_07439	6.90	9.87
g12284	hypothetical protein PV04_01423	5.84	8.62
g12494	exo-beta- -glucanase	-0.70	6.72
g12511	regulator of g signaling superfamily	5.18	7.67
g12629	---NA---	8.90	5.52
g12635	transcription elongation factor S-II	5.38	-1.74
g12849	NB-ARC and TPR domain-containing	6.84	2.21
g12945	NAD(P)-binding Rossmann-fold containing	5.31	-0.42
g12965	hypothetical protein FE78DRAFT_72758	6.41	2.36
g12994	predicted protein	1.91	8.23
g13135	pisatin demethylase	5.98	8.23
g13137	Cytochrome P450	4.68	7.36
g13174	uncharacterized protein FPRO_04301	1.51	6.71
g13196	RNI-like protein	5.37	-0.94
g13209	alpha beta-hydrolase	6.90	9.43
g13377	MFS transporter	6.99	3.01
g13472	hypothetical protein K441DRAFT_7544	2.09	6.62
g13486	hypothetical protein PV07_12346	4.11	9.23
g13487	beta-ig-h3 fasciclin	1.91	8.56
g13492	---NA---	6.28	-0.16
g13508	cell division control 42	8.32	0.90
g13509	Pkinase-domain-containing protein	5.93	-1.74
g13652	---NA---	3.95	-3.90
g13656	hypothetical protein MPH_11376	5.84	-0.16
g13695	uncharacterized protein PAC_01586	8.92	5.28
g13723	hypothetical protein SISNIDRAFT_484147	-0.99	6.11
g13734	hypothetical protein K432DRAFT_290587	4.87	7.99
g13764	PTR2-domain-containing protein	5.96	7.92
	hypothetical protein		
g13789	AOQ84DRAFT_329710	4.50	7.46
g13818	Aspartate--tRNA ligase, cytoplasmic	8.14	4.47
g13820	hypothetical protein W97_07655	6.27	-0.16
g13837	hypothetical protein M413DRAFT_32936	7.50	2.66
	High-affinity methionine permease		
g13841	P50276	4.60	9.06

g13957	hypothetical protein SMAC_02332	5.58	-0.04
g13967	---NA---	6.93	1.81
g13968	RNA-binding protein	6.72	2.28
g14007	hypothetical protein AZE42_07563	7.01	3.91
g14031	hypothetical protein TCE0_034r10074	4.36	-2.16
g14173	hypothetical protein SS1G_01102	8.12	4.24
	hypothetical protein		
g14275	AOQ84DRAFT_380795	3.76	8.02
g14452	hypothetical protein HYPsudRAFT_36721	6.08	-3.90
g14467	hypothetical protein MBM_00672	5.46	-0.94
g14560	hypothetical protein PGRI_095850	6.85	2.34
g14561	---NA---	7.44	2.75
g14717	---NA---	6.85	2.78
g14736	hypothetical protein Hypma_11659	0.90	5.97
g14766	hypothetical protein HCAG_06438	2.89	6.75
g14945	hypothetical protein EPUS_07439	6.39	8.81
g14989	hypothetical protein PV09_01633	6.13	8.50
g15072	hypothetical protein PTTG_27154	-2.07	6.43
	benzoate 4-monooxygenase cytochrome		
g15153	P450	6.21	1.75
g15430	---NA---	6.79	1.06
g15938	hypothetical protein ASPGLDRAFT_22819	7.24	2.06
g15942	hypothetical protein EPUS_00302	-0.99	4.49
g16200	hypothetical protein X797_010364	6.75	9.37
g16235	hypothetical protein ASPZODRAFT_25768	-1.57	5.90
g16255	Sporulation-specific 5	5.68	0.65
g16287	---NA---	3.94	7.35

Table S22: List of 183 DEGs between R (au–10) vs. TW (au–10), and the 65-background variation DEGs are highlighted in red.

Gene id.	Description	TW (au-2010) (log2 RPM)	R (au-2010) (log2 RPM)
g239	Dehydroquinate synthase	7.97	5.24
g548	Choline transport	8.45	6.22
g777	hypothetical protein V490_00898	4.85	7.07
g881	hypothetical protein K432DRAFT_174702	6.24	8.19
g965	---NA---	7.43	3.74
g967	glycosyltransferase family 2	6.88	4.43
g968	hypothetical protein EPUS_03328	6.43	2.37
g1044	vanillyl alcohol oxidase	6.99	2.59
g1045	amp-binding enzyme	9.00	6.50
g1049	hypothetical protein COCC4DRAFT_37502	6.36	2.39
g1050	2og-fe oxygenase family oxidoreductase	7.23	3.07
g1051	hypothetical protein HIM_05906	7.82	3.35
g1052	ABC transporter	10.39	5.63
g1316	-ribulosamine 3- chloroplastic	-1.44	6.89
g1319	hypothetical protein MGYG_05382	4.28	6.42
g1670	hypothetical protein IQ07DRAFT_590219	6.54	1.42
g1816	murein transglycosylase	9.11	4.68
g1817	vtc domain-containing	6.05	1.85
g1825	hypothetical protein AOQ84DRAFT_267913, partial	6.31	2.40
g1947	c2 domain	8.17	10.07
g2497	conserved hypothetical protein	6.72	2.68
g2498	hypothetical protein EPUS_03937	7.85	5.28
g2502	predicted protein	6.41	3.11
g2703	---NA---	4.52	-1.90
g3172	hypothetical protein V500_00208	6.50	3.65

g3722	hypothetical protein EPUS_06199	2.75	6.41
g4104	tyrosine- phosphatase non-receptor type 6	4.97	7.12
g4166	hypothetical protein AOQ84DRAFT_321579	7.68	4.73
g4217	[Laccaria bicolor S238N-H82] XP_001878005	8.02	5.34
g4273	---NA---	7.63	2.05
g4494	hypothetical protein G647_03136	10.81	8.20
g4607	hypothetical protein OIADMADRAFT_35369	1.14	6.30
g4661	hypothetical protein EPUS_00425	2.69	6.72
g4884	hypothetical protein ANI_1_1854104	5.08	8.62
g4963	hypothetical protein FOXB_11579	8.66	4.58
g4995	hypothetical protein L228DRAFT_240605 probable APG9-integral membrane required for Cvt and autophagy transport	7.81	3.96
g5051		5.62	1.01
g5832	carbonic anhydrase	8.46	4.76
g5941	uncharacterized protein PAC_09204	5.73	8.32
g5967	hypothetical protein AZE42_02734	4.94	1.49
g6019	DUF341 domain	6.25	2.37
g6020	reducing type I polyketide synthase	8.68	5.44
g6089	hypothetical protein BAUCODRAFT_146537	4.19	-0.90
g6220	hypothetical protein RHOBADRAFT_67013	6.50	3.02
g6221	hypothetical protein ARAM_006588	5.64	1.01
g6246	Tetratricopeptide-like helical	2.36	6.21
g6296	---NA---	6.50	1.77
g6297	hypothetical protein AOQ84DRAFT_391328	1.08	5.59
g6505	---NA---	6.17	2.40
g6576	benzoate 4-monooxygenase cytochrome P450	8.56	4.95
g6750	---NA---	5.02	-1.09
g7085	TPA_exp: Tyrosinase	6.63	2.19
g7131	---NA---	8.56	5.76
g7265	Collagen alpha-5(VI) chain	7.85	5.56
g7524	hypothetical protein MBM_00672	-1.44	4.42
g7548	---NA---	4.03	-1.32
g7596	---NA---	4.17	-1.32
g7724	AFLR	4.72	-0.44

g7752	hypothetical protein FISHEDRAFT_24821, partial	0.73	5.30
g7839	hypothetical protein K441DRAFT_742369	4.77	-1.32
g7881	hypothetical protein PV08_05142	7.85	4.79
g8030	factor for adipocyte	2.51	6.76
g8218	outer membrane autotransporter barrel	7.84	3.76
g8261	low-affinity ammonium transporter	8.02	6.22
g8535	hydroxyindole o	9.34	7.59
g8548	hypothetical protein PV09_00215	8.61	5.47
g8781	multicopper oxidase	2.69	6.35
g8782	---NA---	4.09	-1.32
g8882	hypothetical protein OIADMADRAFT_105967	7.59	3.89
g8994	hypothetical protein SCHPADRAFT_996133	5.70	0.80
g8995	E3 ubiquitin ligase complex SCF subunit scon	5.48	0.49
g9086	terpene cyclase	5.98	1.01
g9319	Podospira anserina S mat genomic DNA chromosome supercontig 2	7.33	4.04
g9558	---NA---	-3.60	4.00
g9882	amine flavin-containing superfamily	4.84	7.37
g10181	hypothetical protein AK830_g408	5.87	7.83
g10296	hypothetical protein AOQ84DRAFT_351571	5.49	8.31
g10332	hypothetical protein PV08_05721	5.47	7.42
g10528	hypothetical protein EPUS_00425	2.58	7.06
g10672	hypothetical protein A103_04947	0.88	5.41
g10721	catalase	5.43	7.62
g10814	---NA---	9.12	5.54
g10851	---NA---	7.76	4.85
g10882	DEAD DEAH box helicase	6.59	8.80
g10948	---NA---	5.38	1.65
g10992	---NA---	6.96	4.12
g11004	galectin-like protein	10.62	8.49
g11384	nb-arc and tpr domain	4.22	6.50
g11436	related to multidrug resistant	-2.44	5.42
g11695	heterokaryon incompatibility	3.44	6.33
g11749	hypothetical protein TSTA_020850	-3.60	3.30

g11900	histidine kinase	6.30	1.71
g11970	Vegetative incompatibility HET-E-1	3.34	6.57
g12020	hypothetical protein LY89DRAFT_747417	7.24	4.98
g12021	uncharacterized protein PAC_08291	8.45	5.05
g12022	kinase-like domain	2.34	5.82
g12023	hypothetical protein CPC735_009070	1.36	5.99
g12045	serine threonine kinase	6.05	1.46
g12046	hypothetical protein BATDEDRAFT_24227	5.48	-0.58
g12077	---NA---	5.89	-0.90
g12122	---NA---	2.02	6.77
g12125	---NA---	-0.64	4.39
g12126	---NA---	-0.12	6.23
g12169	flavin-binding monooxygenase	3.49	6.47
g12170	1, paxillin biosynthesis enzyme PAXC	4.89	7.92
g12197	gag-pol poly	0.36	4.68
g12261	hypothetical protein EPUS_07439	7.75	9.71
g12460	---NA---	4.43	-0.44
g12505	solid-state culture specific	2.36	7.03
g12696	Cytochrome P450 pisatin demethylase	-1.44	4.30
g12929	putative subtilase	6.50	2.76
g12994	predicted protein	5.55	8.07
g13009	isoform Er16 of ankyrin-1	-3.60	4.12
g13010	serine threonine kinase	-3.60	6.08
g13011	serine threonine kinase	-3.60	4.40
g13012	hypothetical protein EPUS_03941	-0.86	7.12
g13060	laccase-like multicopper oxidase	8.56	10.83
g13121	serine threonine- kinase chk1	-3.60	4.18
g13122	---NA---	1.20	6.36
g13147	related to ferroportin 1	5.84	1.01
g13197	hypothetical protein W97_05036	3.43	6.55
g13209	alpha beta-hydrolase	5.48	9.28
g13275	---NA---	3.97	-1.90
g13337	Calcium calmodulin-dependent kinase type	-0.44	4.59
g13376	hypothetical protein sscl_09g068980	6.30	0.96

g13377	MFS transporter	9.17	2.85
	Sexual differentiation process subtilase-type		
g13460	ase isp6	4.03	-1.32
g13476	hypothetical protein PV11_09631	9.44	6.77
g13494	---NA---	-1.86	3.80
g13508	cell division control 42	6.03	0.74
g13593	hypothetical protein LY89DRAFT_609882	5.18	0.74
g13841	High-affinity methionine permease P50276	4.91	8.90
g13876	Ferritin-like domain	4.97	7.27
g14007	hypothetical protein AZE42_07563	-3.44	3.76
g14089	---NA---	4.68	1.10
g14320	hypothetical protein ALT_0985	6.65	9.46
	S-adenosyl-L-methionine-dependent		
g14365	methyltransferase	-3.60	3.64
g14382	hypothetical protein CTHT_0044650	-0.27	5.90
g14424	High-affinity methionine permease P50276	1.64	6.35
g14467	hypothetical protein MBM_00672	4.36	-1.09
g14468	---NA---	7.42	5.49
g14525	P30 adhesin	7.33	5.13
g14540	Lysozyme	7.99	5.22
g14650	hypothetical protein V499_07880	-1.12	6.32
	ATPase of HSP90 chaperone DNA		
g14656	topoisomerase II histidine kinase	6.61	8.67
g14707	NAD(P)-binding protein	-3.60	5.71
g14708	hypothetical protein CC84DRAFT_1195088	-3.60	3.91
g14709	hypothetical protein H101_03049	-3.60	4.77
g14736	hypothetical protein Hypma_11659	1.02	5.81
g14766	hypothetical protein HCAG_06438	4.01	6.60
g14803	---NA---	5.29	0.01
g14896	hypothetical protein SS1G_06147	0.56	6.36
g14910	hypothetical protein BAUCODRAFT_145551	4.14	7.46
g14930	Ubiquitin carboxyl-terminal hydrolase	9.72	6.10
	hypothetical protein A1Q2_08520		
g14931	(mitochondrion)	8.95	5.88
g14932	hydrolase, partial	9.80	6.62

g14933	hypothetical protein CONCODRAFT_139759	9.37	6.11
g14951	hypothetical protein AOQ84DRAFT_306480, partial	-3.60	3.62
g14990	hypothetical protein OI DMADRAFT_178813	0.88	6.33
g15008	---NA---	-3.44	3.88
g15061	hypothetical protein AGABI2DRAFT_176739	3.43	7.49
g15072	hypothetical protein PTTG_27154	-0.64	6.27
g15160	---NA---	7.24	3.19
g15178	hypothetical protein PV05_09409	2.60	8.42
g15263	hypothetical protein MGYG_02417	2.02	6.56
g15272	hypothetical protein ASPVEDRAFT_41603, partial	-3.60	4.39
g15422	Cytochrome P450	-3.60	4.26
g15423	cytochrome P450	-3.60	5.36
g15432	hypothetical protein EPUS_06399	0.36	5.00
g15626	hypothetical protein K432DRAFT_430842	-0.64	4.32
g15640	hypothetical protein PTT_10076	7.62	5.14
g15641	---NA---	9.72	6.93
g15642	permease, partial	8.33	5.60
g15647	hypothetical protein K432DRAFT_430842	-2.44	4.05
g15656	benzoate 4-monooxygenase cytochrome P450	-3.60	6.02
g15744	hypothetical protein BC1G_01702	2.28	6.17
g15831	hypothetical protein VE03_06097	0.56	5.47
g15895	hypothetical protein AOL_s00215g640	-1.12	4.50
g16168	hypothetical protein L228DRAFT_242344	9.45	6.99
g16231	---NA---	7.87	4.29
g16262	hypothetical protein CORT_OF02850	5.55	2.23
g16274	---NA---	6.15	1.27
g16399	---NA---	7.46	4.56

Table S23: List of 84 DEGs between A (au-09) vs. TB (su-12), and the 11-background variation DEGs are highlighted in red.

Gene id.	Description	A (au-2009) (log2 RPM)	TB (su-2012) (log2 RPM)
g239	Dehydroquinase synthase	4.88	7.86
g246	---NA---	8.25	1.72
g581	ABC transporter	7.04	2.77
g689	---NA---	7.16	1.53
g816	hypothetical protein TGAM01_04190	6.89	3.27
g1005	flocculation suppression	4.07	6.81
g1006	hypothetical protein VF21_09233	4.14	7.32
g1045	amp-binding enzyme	4.18	6.90
g1050	2og-fe oxygenase family oxidoreductase FAD dependent oxidoreductase	2.78	7.09
g1256	superfamily	7.59	4.90
g1316	-ribulosamine 3- chloroplastic	6.87	-2.14
g1639	Piwi-domain-containing protein	6.94	0.67
g2053	---NA---	8.62	4.69
g2113	hypothetical protein W97_03127	7.62	0.81
g2749	hypothetical protein BAUCODRAFT_27901	6.35	-0.25
g3453	hypothetical protein M422DRAFT_265232 [Laccaria bicolor S238N-H82]	7.03	1.03
g4217	XP_001878005	4.48	8.62
g4444	hypothetical protein A103_09169	6.71	-1.91
g4480	hypothetical protein SS1G_00468	7.56	2.27
g4604	---NA---	6.77	2.62
g4718	---NA---	6.24	0.28
g5248	hypothetical protein PTT_05645	6.54	-1.40
g5423	hypothetical protein MAJ_10788, partial	6.99	3.93
g5452	---NA---	7.75	2.35
g5744	---NA---	7.67	-0.16
g6026	hypothetical protein FE78DRAFT_72758	7.34	1.77
g6077	hypothetical protein MGG_02083	7.25	1.57
g6504	---NA---	3.93	-4.86
g6590	hypothetical protein HMPREF1541_09740	6.19	8.44
g6711	carbohydrate-binding module family 13	-1.01	7.67
g6929	---NA---	7.10	2.45

g7123	MFS general substrate transporter	8.32	4.82
g7358	hypothetical protein A109_02403	3.74	7.08
g7418	---NA---	6.62	-2.11
	hypothetical protein		
g7747	BOTBODRAFT_118501	7.23	1.86
g7881	hypothetical protein PV08_05142	5.78	8.40
g9151	hypothetical protein H101_03408	7.15	2.57
g9171	sterol desaturase family	6.73	-0.35
g9390	---NA---	7.06	-0.33
g9449	---NA---	7.87	2.53
g9517	scp-like extracellular	7.76	3.45
g9751	mate family transporter	4.99	7.27
g9822	ATP-dependent RNA helicase ded1,	7.60	-3.64
g10277	retinol dehydrogenase 12	5.57	8.52
g10792	---NA---	8.24	6.14
g10793	p53-like transcription factor	8.07	5.77
g10851	---NA---	5.54	9.86
g10880	---NA---	7.49	-2.31
g11004	galectin-like protein	3.31	8.43
g11327	beta-lactamase transpeptidase	5.69	-5.73
g11331	O-methyltransferase	3.07	7.64
g11415	hypothetical protein FE78DRAFT_72758	6.85	1.16
g11615	heat shock 30	7.49	4.59
g12222	FAS1 domain-containing	3.79	7.80
g12615	---NA---	7.98	10.73
g12965	hypothetical protein FE78DRAFT_72758	7.62	3.14
g12994	predicted protein	2.84	8.13
g13209	alpha beta-hydrolase	4.44	7.43
g13317	hypothetical protein EPUS_06272	7.29	3.39
g13492	---NA---	6.29	0.73
g13508	cell division control 42	6.21	-4.90
g13757	TPR domain-containing	3.72	-4.97
g13820	hypothetical protein W97_07655	6.92	1.67
g13837	hypothetical protein M413DRAFT_32936	8.05	1.57
g13967	---NA---	7.22	4.42
g13968	RNA-binding protein	7.37	4.95
g14225	hypothetical protein OGAPODRAFT_12856	6.85	1.80
g14474	---NA---	4.88	8.03
g14555	hypothetical protein M422DRAFT_39030	3.64	6.71
g14560	hypothetical protein PGRI_095850	6.61	2.41
g14629	hypothetical protein W97_08269	5.75	7.81

g14717	---NA---	6.83	2.77
g14762	hypothetical protein K441DRAFT_650523	4.81	7.85
g14998	---NA---	7.49	-0.84
g15231	---NA---	5.49	7.82
g15278	---NA---	7.01	1.98
g15430	---NA---	7.34	0.23
g15614	GroES-like protein	3.78	8.05
g15651	hydrolase, partial	6.62	2.56
g15670	hydrolase, partial	7.26	3.69
g15914	---NA---	5.90	0.96
g15938	hypothetical protein ASPGLDRAFT_22819	6.56	-0.39
g16255	Sporulation-specific 5	6.93	3.34
g16309	permease, partial	7.11	2.93

Table S24: List of 294 DEGs between R (su-12) vs. TB (su-12), and the 63-background variation DEGs are highlighted in red.

Gene id.	Description	R (Su-2012) (log2 RPM)	TB (Su-2012) (log2 RPM)
g199	hypothetical protein ZTR_09325	7.70	5.67
g238	catechol o-methyltransferase	5.08	7.37
g239	Dehydroquinase synthase	7.25	9.43
g241	multidrug transporter	5.66	3.41
g271	class III chitinase 1	4.03	5.95
g272	glycoside hydrolase family 18 , partial	5.05	6.78
g363	carboxypeptidase S1, related to vacuolar sorting VPS1, dynamin, and	7.08	8.85
g415	related s	5.06	6.95
g442	hypothetical protein LY89DRAFT_716349	7.71	5.83
g489	---NA---	3.18	5.66
g505	geranylgeranyl pyrophosphate synthetase conidial pigment biosynthesis oxidase Abr1	6.24	7.59
g530	brown 1	9.43	6.99
g534	aromatic peroxygenase	5.67	7.44
g548	Choline transport	7.09	9.04
g574	hypothetical protein ASPCAL04919	8.07	6.98
g967	glycosyltransferase family 2	4.89	7.49
g1001	udp-glucuronosyltransferase 2a3	5.18	7.67
g1005	flocculation suppression	4.83	7.01
g1006	hypothetical protein VF21_09233	5.38	8.54
g1044	vanillyl alcohol oxidase	3.97	7.16
g1045	amp-binding enzyme	7.51	10.54
g1049	hypothetical protein COCC4DRAFT_37502	3.44	7.13
g1050	2og-fe oxygenase family oxidoreductase	4.16	7.28
g1051	hypothetical protein HIM_05906	4.55	8.18
g1052	ABC transporter	7.09	10.69
g1053	---NA---	-0.08	3.26
g1145	hypothetical protein NFIA_105120	2.32	4.76
g1155	polysaccharide deacetylase family	-0.76	4.88
g1211	hypothetical protein CONPUDRAFT_165241	3.89	6.50
g1215	Serine ase 2	2.98	5.91

g1263	subtilisin-like protease	-2.11	2.13
g1322	hypothetical protein ANI_1_2290184	3.21	8.00
g1670	hypothetical protein IQ07DRAFT_590219	1.62	4.13
g1753	---NA---	1.70	3.92
g1802	carbohydrate-binding module family 18	9.90	11.24
g1816	murein transglycosylase	5.95	8.51
g1817	vtc domain-containing hypothetical protein AOQ84DRAFT_267913,	2.36	5.87
g1825	partial	2.87	5.58
g2129	reverse transcriptase,	8.33	6.95
g2223	Uncharacterized protein LW93_9556	4.22	6.15
g2497	conserved hypothetical protein	3.76	7.31
g2498	hypothetical protein EPUS_03937	6.42	7.81
g2502	predicted protein	2.39	5.96
g2503	hypothetical protein EV44_g1971	-0.74	3.51
g2504	hypothetical protein ASPZODRAFT_1392612 A5178_ARTBC ame: Full=Uncharacterized	0.27	3.07
g2698	secreted ARB_05178 Flags: Precursor	5.86	3.47
g2739	---NA---	9.07	10.40
g2878	uncharacterized protein PAC_18549	4.29	6.11
g2882	C6 transcription factor	4.54	6.49
g3143	carboxypeptidase S1	8.53	10.06
g3168	amino acid transporter	7.19	8.59
g3172	hypothetical protein V500_00208	2.89	6.05
g3190	potassium uptake	10.75	8.92
g3485	hypothetical protein BC1G_05584	7.14	8.83
g3506	hypothetical protein EPUS_02122	5.66	8.27
g3540	hypothetical protein GYMLUDRAFT_706765	6.95	4.47
g3838	---NA---	7.29	6.09
g3893	hypothetical protein GLAREA_08855	7.50	4.90
g4056	---NA---	-1.21	3.55
g4166	hypothetical protein AOQ84DRAFT_321579	4.05	6.74
g4206	hypothetical protein FOIG_10308	0.77	3.59
g4217	[Laccaria bicolor S238N-H82] XP_001878005	6.11	8.15
g4239	aat1p	-1.82	3.97
g4273	---NA---	1.17	4.60
g4322	---NA---	-0.74	3.09
g4508	Scytalone dehydratase	7.82	5.97
g4509	1,3,6,8-tetra-HN reductase	9.42	6.00
g4510	transcription factor	8.38	6.40

g4511	non-reducing type I polyketide synthase	12.23	10.04
g4607	hypothetical protein OI DMADRAFT_35369	7.54	3.66
g4651	PA14 domain	0.55	4.15
g4652	toxin subunit	-2.53	2.33
g4653	toxin subunit , partial	3.56	6.57
g4702	Cytochrome P450	1.24	3.84
g4703	hypothetical protein FOVG_18288	2.10	4.59
g4922	hypothetical protein FVEG_12244	1.78	4.77
g4963	hypothetical protein FOXB_11579	3.57	5.95
g4995	hypothetical protein L228DRAFT_240605	2.37	6.16
g5147	hypothetical protein K432DRAFT_397027	10.30	8.84
g5155	Uncharacterized transporter	4.56	7.97
g5156	hypothetical protein Z518_02103	0.66	4.10
g5157	MOSC domain containing	5.25	7.67
g5284	DNA mismatch repair , core	2.75	5.84
g5388	transcription initiation factor iif	3.87	5.41
g5401	Translation initiation factor IF-2	-1.11	2.52
g5410	hypothetical protein VE04_07924	9.43	7.94
g5476	hypothetical protein AOQ84DRAFT_431293	4.30	7.08
g5477	hypothetical protein AGABI2DRAFT_123004	0.90	3.53
g5832	carbonic anhydrase	5.00	6.89
g5967	hypothetical protein AZE42_02734	1.36	4.13
g6018	major facilitator superfamily transporter	5.36	7.73
g6019	DUF341 domain	2.94	6.29
g6020	reducing type I polyketide synthase	5.65	8.34
g6021	reducing type I polyketide synthase	0.32	3.84
g6022	reducing type I polyketide synthase	1.15	4.00
g6220	hypothetical protein RHOBADRAFT_67013	2.89	5.85
g6221	hypothetical protein ARAM_006588	-0.03	3.25
g6296	---NA---	4.03	7.01
g6386	hypothetical protein V490_07430	3.28	6.36
	S-adenosyl-L-methionine-dependent		
g6501	methyltransferase	4.63	8.03
g6505	---NA---	0.99	5.03
g6576	benzoate 4-monooxygenase cytochrome P450	4.05	6.27
g6633	[Laccaria bicolor S238N-H82] XP_001878005	-2.69	4.64
g6711	carbohydrate-binding module family 13	3.75	6.64
g6747	barwin-like endoglucanase	2.66	4.96
g6756	---NA---	-0.24	3.37
g6930	hypothetical protein DOTSEDRAFT_138359	-1.37	2.42

	related to calcium-independent phospholipase		
g6933	A2	7.26	5.11
g7085	TPA_exp: Tyrosinase	2.03	4.46
g7157	hypothetical protein ALT_8794	3.05	5.30
g7265	Collagen alpha-5(VI) chain	5.80	7.69
g7331	LAFE_0E15038g1_1	8.83	7.34
g7359	Polysaccharide synthase Cps1	5.53	7.67
g7528	hypothetical protein ASPWEDRAFT_144002	0.03	3.46
g7530	hypothetical protein LY89DRAFT_743438	1.68	4.63
g7641	complexed with Cdc5 Cwf16	4.58	6.37
g7654	hypothetical protein L228DRAFT_267680	5.40	7.50
g7681	hypothetical protein H109_01118	3.84	0.74
g7724	AFLR	1.61	4.66
g7839	hypothetical protein K441DRAFT_742369	-1.32	3.36
g8011	cysteine dioxygenase Cdo1	2.67	4.84
g8039	c6 zinc finger domain containing	6.27	7.81
g8089	---NA---	5.82	7.41
g8179	Collagen triple helix repeat-containing	3.33	-0.70
g8190	SGNH hydrolase	3.45	-1.32
g8218	outer membrane autotransporter barrel	6.07	8.20
g8246	hypothetical protein COCSADRAFT_140624	7.36	5.78
g8406	dipeptidyl-peptidase III	5.39	6.61
g8535	hydroxyindole o	7.13	8.78
g8548	hypothetical protein PV09_00215	6.02	8.33
g8579	hypothetical protein V496_03516	2.64	4.94
g8581	1,3-beta glucanase	5.84	7.23
g8700	small oligopeptide OPT family	5.72	8.60
g8765	hypothetical protein	6.31	3.36
g8781	multicopper oxidase	7.42	5.22
g8782	---NA---	-1.24	2.84
g8838	SGNH hydrolase	6.77	4.10
g8882	hypothetical protein OIADMADRAFT_105967	4.01	7.31
g8994	hypothetical protein SCHPADRAFT_996133	0.71	4.15
g8995	E3 ubiquitin ligase complex SCF subunit scon	-0.58	4.22
g9075	hypothetical protein L228DRAFT_263014	4.12	6.17
g9082	cytochrome P450 monooxygenase	0.59	4.71
g9085	cytochrome P450	1.74	6.39
g9086	terpene cyclase	-0.80	3.70
g9099	---NA---	-2.69	1.72
g9159	hypothetical protein HYPBUDRAFT_152817	1.32	3.74

g9160	RNA-binding domain-containing	2.07	4.69
g9289	hypothetical protein TDEL_OA05340	3.82	5.99
	Podospira anserina S mat genomic DNA		
g9319	chromosome supercontig 2	4.28	7.33
g9320	hypothetical protein XA68_103	1.85	5.12
g9346	Pre-rRNA-processing PNO1	2.39	5.02
g9882	amine flavin-containing superfamily	7.97	6.60
g9921	gaba permease	8.07	9.15
g9938	quino amine dehydrogenase beta chain	8.07	6.90
g10178	---NA---	-0.45	4.36
g10187	hypothetical protein EPUS_00229	8.52	7.44
g10252	polyketide synthase	-0.01	3.30
g10296	hypothetical protein AOQ84DRAFT_351571	7.36	4.02
g10304	hypothetical protein CC84DRAFT_1102722	0.66	4.28
	NADPH-dependent 1-acyldihydroxyacetone		
g10305	phosphate reductase	5.12	7.46
g10306	reducing type I polyketide synthase 10	4.75	7.43
g10547	hypothetical protein FG05_01788	5.91	4.09
g10555	hypothetical protein EURHEDRAFT_526790	5.93	4.29
g10570	UPF0255 protein	6.52	5.06
g10725	UPF0061-domain-containing protein	-2.69	2.40
g10882	DEAD DEAH box helicase	7.79	5.98
g10919	Uncharacterized lipo ybbD P40406	1.93	5.39
g10948	---NA---	-2.69	2.49
g10973	betaine lipid synthase	6.54	7.95
g10992	---NA---	3.81	6.30
g11004	galectin-like protein	8.17	10.20
g11138	Common central domain of tyrosinase	8.66	6.08
g11144	kinase-like domain	-0.74	2.17
g11184	transitional endoplasmic reticulum ATPase	0.76	4.21
g11247	hypothetical protein K441DRAFT_659689	7.00	5.22
g11259	MFS multidrug resistance transporter,	8.63	6.82
g11427	hypothetical protein PHSY_002683	2.49	4.74
g11434	uncharacterized protein PAC_17075	5.90	-2.03
g11435	uncharacterized protein PAC_10069	8.71	2.95
g11436	related to multidrug resistant	6.94	2.23
g11437	uncharacterized protein PAC_17077	6.31	1.87
g11468	uncharacterized protein CPUR_05377	2.73	-0.37
g11482	Metal regulatory transcription factor 1	-0.53	4.03
g11483	cytochrome p450	2.60	6.79

g11485	hypothetical protein BofuT4_P025320.1	4.56	7.17
g11486	cytochrome p450	2.26	5.77
g11622	hypothetical protein V498_10729	6.22	3.88
g11627	hypothetical protein EPUS_00425	4.99	1.97
g11695	heterokaryon incompatibility	6.08	2.69
g11711	HC-toxin efflux carrier TOXA	9.23	6.23
g11712	methyltransferase	7.72	1.93
g11713	hypothetical protein M437DRAFT_59426	8.99	3.79
g11797	NAD(P)-binding protein	4.06	6.06
g11900	histidine kinase	0.82	4.63
	MFS transporter, ACS family, allantate		
g11954	permease	2.69	5.90
g11971	cytochrome P450	0.70	4.24
g12020	hypothetical protein LY89DRAFT_747417	5.52	8.10
g12021	uncharacterized protein PAC_08291	5.00	7.72
g12045	serine threonine kinase	1.17	4.38
g12046	hypothetical protein BATDEDRAFT_24227	-0.71	4.22
g12048	hypothetical protein XA68_2246	1.93	5.36
g12049	fructose-bisphosphate aldolase,	2.61	5.59
g12050	---NA---	2.28	5.21
g12052	---NA---	2.24	4.95
g12077	---NA---	0.47	5.57
g12111	F5 8 type C domain-containing	1.92	4.41
g12120	hypothetical protein A109_01179	4.09	1.36
g12122	---NA---	6.47	3.16
g12125	---NA---	5.91	4.29
g12126	---NA---	7.17	5.30
g12306	hypothetical protein M404DRAFT_122096	-1.03	3.71
g12398	benzoate 4-monooxygenase cytochrome p450	5.59	2.70
g12411	hypothetical protein V494_07452	0.76	3.65
g12432	---NA---	-1.82	2.91
g12460	---NA---	-0.08	4.07
g12505	solid-state culture specific	6.86	4.32
g12659	Sterol regulatory element-binding 1	8.80	7.32
g12771	amino acid transporter	5.67	7.57
g12852	predicted protein	1.32	4.65
g12859	---NA---	-0.74	2.46
g12929	putative subtilase	1.55	4.70
g12974	Isotrichodermin C-15 hydroxylase	6.12	7.48
g12991	uncharacterized protein FFUJ_00300	6.34	2.38

g12999	integral membrane	2.79	6.66
g13045	---NA---	3.17	5.84
g13046	alpha beta-Hydrolase	6.76	8.45
g13127	---NA---	-0.45	3.29
g13147	related to ferroportin 1	1.65	5.24
g13150	hypothetical protein ASPCAL07495	-2.69	3.47
g13174	uncharacterized protein FPRO_04301	7.35	5.97
g13255	---NA---	4.17	6.32
g13256	hypothetical protein ASPGLDRAFT_182435	7.46	3.36
g13257	catalytic	10.06	6.67
g13258	C6 zinc finger domain	5.96	3.44
g13376	hypothetical protein sscl_09g068980	-0.16	4.08
g13377	MFS transporter	3.20	7.39
g13424	thermosensitive gluconokinase	7.39	5.83
g13440	carbohydrate-binding module family 32	6.48	4.11
	Sexual differentiation process subtilase-type		
g13460	ase isp6	0.13	3.30
g13486	hypothetical protein PV07_12346	9.26	5.06
g13487	beta-ig-h3 fasciclin	8.82	5.23
g13511	cysteine ase	5.80	7.38
g13593	hypothetical protein LY89DRAFT_609882	0.29	4.07
g13713	hypothetical protein EPUS_05087	6.37	4.39
g13734	hypothetical protein K432DRAFT_290587	8.35	4.27
g13737	---NA---	1.05	3.64
g13782	photosystem I core	0.05	3.91
g13783	---NA---	-1.03	3.28
g13786	gaba permease	6.88	4.86
g13788	hypothetical protein VF21_00684	5.01	-0.11
g13789	hypothetical protein AOQ84DRAFT_329710	6.30	2.21
g13841	High-affinity methionine permease P50276	9.42	6.29
g13917	hypothetical protein PIIN_06735	2.84	5.50
g13967	---NA---	1.59	4.16
g13968	RNA-binding protein	2.37	4.34
g14007	hypothetical protein AZE42_07563	3.81	5.95
g14086	---NA---	0.95	5.07
g14248	---NA---	0.16	4.69
g14275	hypothetical protein AOQ84DRAFT_380795	7.37	3.95
g14364	PenC	3.91	-0.35
	S-adenosyl-L-methionine-dependent		
g14365	methyltransferase	4.77	-1.07

g14467	hypothetical protein MBM_00672	1.32	4.56
g14516	hypothetical protein VE03_08214	5.15	6.84
g14540	Lysozyme	4.59	6.74
g14603	predicted protein	-1.03	3.99
g14731	hypothetical protein BofuT4_P088950.1	6.19	7.28
g14762	hypothetical protein K441DRAFT_650523	5.03	6.87
g14803	---NA---	1.00	5.16
g14930	Ubiquitin carboxyl-terminal hydrolase	6.48	8.15
g14931	hypothetical protein A1Q2_08520 (mitochondrion)	6.24	8.13
g14932	hydrolase, partial	6.69	8.91
g14933	hypothetical protein CONCODRAFT_139759	6.24	8.52
g14956	---NA---	0.45	5.04
g14964	---NA---	3.37	5.74
g15072	hypothetical protein PTTG_27154	7.99	4.22
g15092	hypothetical protein PFICI_07720	3.56	6.32
g15360	---NA---	5.66	3.34
g15421	sporulation associated	2.84	5.09
g15501	hypothetical protein PFICI_15058	-1.45	3.56
g15502	lectin-2	0.90	5.24
g15539	Clavaminate synthase	3.43	5.50
g15641	---NA---	6.66	8.58
g15642	permease, partial	5.62	7.47
g15680	hypothetical protein ASPWEDRAFT_28951	2.44	-0.92
g15702	hypothetical protein I308_02452	-1.11	2.98
g15707	conserved hypothetical protein	6.06	4.63
g15759	integral membrane	2.31	5.35
g15995	MFS transporter	2.48	5.70
g16044	polyketide synthase	-0.87	3.81
g16045	hypothetical protein DHEL01_09803	2.30	5.54
g16121	hypothetical protein, partial	5.48	2.76
g16231	---NA---	5.55	7.67
g16262	hypothetical protein CORT_0F02850	0.27	5.56
g16282	mfs multidrug transporter	4.99	7.47

Table S25: List of 317 DEGs between TW (au–09) vs. TW (au–10) vs. TW (Su–11) vs. TW (su–12), and the 105-background variation DEGs are highlighted in red.

Gene id.	Description	TW (au-2009) (log2 RPM)	TW (au-2010) (log2 RPM)	TW (su-2011) (log2 RPM)	TW (su-2012) (log2 RPM)
g7	hypothetical protein FOMPIDRAFT_142443	7.41	4.51	7.94	7.35
g99	related to DUF1295 domain	2.70	6.42	5.01	4.54
g105	---NA---	7.53	-1.76	6.93	-0.27
g174	30 kDa heat shock	6.81	3.45	6.81	3.23
g313	uncharacterized protein RSE6_00399	7.02	3.39	6.20	6.58
g365	---NA---	7.94	4.63	8.83	8.79
g366	hypothetical protein Hypma_07800	7.50	3.82	7.20	7.27
g640	deacylase-like protein	7.39	3.90	7.00	7.06
g648	HSP20-like chaperone	7.63	4.60	7.45	7.08
g881	hypothetical protein K432DRAFT_174702	7.79	5.06	8.08	7.63
g882	hypothetical protein BON22_5114	8.25	5.58	8.27	7.88
g1029	hypothetical protein PIIN_06366	7.40	4.60	7.38	7.34
g1045	amp-binding enzyme	7.84	4.94	8.11	8.16
g1315	hypothetical protein O988_04557	6.78	1.72	6.36	6.69
g1316	-ribulosamine 3- chloroplastic	1.16	6.89	6.64	4.12
g1317	aldo keto reductase	6.71	2.09	4.77	4.64
g1319	hypothetical protein MGYG_05382	7.56	4.74	8.30	7.87
g1376	flavin-containing amine oxidasedehydrogenase	7.29	2.12	8.40	8.07

g1424	glycoside hydrolase family 32	6.18	-0.52	6.28	6.77
g1550	hypothetical protein UCDS831_g05414	7.20	1.08	0.27	3.21
g1773	unnamed protein product	5.82	-1.24	-0.67	2.29
g1885	integral membrane	4.39	7.40	5.72	6.91
g1945	meiotically up-regulated	8.26	4.71	7.52	8.08
g1946	meiotically up-regulated	7.36	4.48	7.99	6.89
g1947	c2 domain	6.11	1.32	2.29	1.58
g1964	hypothetical protein W97_04954	7.02	3.32	2.67	1.90
g2075	conserved hypothetical protein	7.17	4.35	7.02	6.67
g2355	hypothetical protein K432DRAFT_18341	7.34	2.80	6.05	7.48
g2495	hypothetical protein OBBRIDRAFT_649165	7.00	4.13	5.03	6.17
g2673	---NA---	7.86	4.62	8.53	8.17
g2782	cyanate hydratase	7.00	2.92	3.37	1.96
g2784	hypothetical protein ASPZODRAFT_19280	6.33	0.42	3.31	1.21
g3136	DUF1264-domain-containing protein [Glonium stellatum]	1.13	6.43	5.12	4.23
g3166	predicted protein	7.15	1.86	5.15	4.85
g3216	hypothetical protein W97_00580	6.77	0.26	4.05	4.46
g3329	---NA---	6.52	0.46	4.67	5.00
g3383	hypothetical protein FPOA_12054	7.30	1.66	4.88	5.34
g3462	hypothetical protein V495_01031	2.58	6.58	6.06	4.27
g3537	---NA---	4.22	-5.46	-1.35	0.34
g3564	hypothetical protein BAUCODRAFT_27538	5.97	-2.01	0.64	2.03
g3703	---NA---	-0.34	6.35	3.23	4.48
g3719	---NA---	7.53	3.23	8.38	8.11
g3722	hypothetical protein EPUS_06199	7.99	4.51	9.85	8.77
g3842	hypothetical protein AOL_s00054g438	7.65	-2.96	8.98	3.96

g3893	hypothetical protein GLAREA_08855	6.38	0.05	6.14	6.34
g4180	related to LEA domain	6.51	1.20	6.17	7.25
g4273	---NA---	3.92	7.46	7.60	6.83
g4496	transmembrane protein UsgS	4.20	7.27	7.42	6.50
g4661	hypothetical protein EPUS_00425	-2.26	6.39	6.17	5.42
g4676	conidiation-specific 6	6.43	-0.20	3.11	3.59
	probable APG9-integral membrane required for Cvt and				
g5051	autophagy transport	6.65	8.83	6.38	6.90
g5073	NADP-dependent alcohol dehydrogenase 6	7.19	-1.29	4.09	3.09
g5128	NAD(P)-binding protein	6.91	-3.53	5.14	2.21
g5168	Subtilisin-like protease	6.72	-1.95	5.19	2.76
g5204	interferon-induced gtp-binding mx1	6.95	-3.68	3.84	1.68
g5491	Sac2 family	7.43	2.05	4.31	4.92
g5542	---NA---	6.39	0.35	4.70	2.98
g5872	---NA---	7.18	3.41	7.35	7.22
g5951	MFS general substrate transporter	7.20	0.95	4.90	5.11
g6028	oxidoreductase	7.39	-4.52	0.22	-3.49
g6036	---NA---	-1.24	6.05	5.28	5.56
g6141	cell division control 15	6.34	0.26	5.13	3.28
g6297	hypothetical protein AOQ84DRAFT_391328	6.65	-2.53	3.10	1.76
g6328	uncharacterized protein RCO7_02887	6.76	-3.40	3.59	1.76
g6341	general substrate transporter	6.72	2.06	8.50	5.42
g6342	---NA---	6.43	-2.70	1.15	-1.67
g6389	---NA---	7.16	-4.32	4.64	1.95
g6391	hypothetical protein LLEC1_03182	-1.18	5.65	5.36	4.77
g6501	S-adenosyl-L-methionine-dependent methyltransferase	4.43	7.57	1.93	2.15

g6599	hypothetical protein ASPWEDRAFT_28951	3.86	7.20	6.86	6.33
g6633	[Laccaria bicolor S238N-H82] XP_001878005	0.55	7.97	1.32	-1.61
g6670	PLP-dependent transferase	-0.90	6.79	6.49	3.50
g6711	carbohydrate-binding module family 13	2.79	7.08	7.21	5.84
g6976	hypothetical protein CY34DRAFT_71314	4.55	6.23	9.65	6.21
g7339	hypothetical protein PAXINDRAFT_97213	6.40	7.00	8.56	8.02
g7358	hypothetical protein A1O9_02403	7.79	5.46	5.13	6.65
g7495	hypothetical protein IQ06DRAFT_80159	2.92	-2.17	-4.03	-2.53
g7524	hypothetical protein MBM_00672	5.08	6.79	7.74	5.26
g7580	hypothetical protein MVEG_03845	1.49	0.17	-4.69	-2.59
g7753	NA	6.53	5.26	4.38	5.99
g7788	hypothetical protein VE01_07399	5.26	6.10	7.42	7.71
g7816	glycoside hydrolase family 5	6.21	4.06	8.19	6.86
g7936	atg7p	2.76	4.23	6.16	5.09
g8007	---NA---	-1.28	5.60	9.26	7.84
g8009	hypothetical protein MVEG_02501	3.10	-1.54	-4.13	-3.65
g8072	disaggregation chaperone	5.46	7.17	8.21	7.41
g8163	TPA: Zn(II)2Cys6 transcription factor (Eurofung)	6.53	4.34	2.50	4.90
g8164	cytochrome P450	3.10	2.49	6.58	1.84
g8206	hypothetical protein L228DRAFT_244052	6.39	7.26	8.54	6.80
g8210	hypothetical protein SS1G_04281	0.18	2.74	5.58	3.16
g8211	---NA---	6.18	5.52	1.69	4.05
g8251	bactericidal permeability-increasing	6.08	4.88	2.21	4.57
g8255	hypothetical protein CIRG_01504	7.56	5.92	4.68	7.15
g8306	---NA---	6.15	3.66	3.27	6.21
g8387	hypothetical protein EPUS_00229	6.14	5.68	2.56	4.38

g8407	---NA---	6.12	7.08	8.35	7.45
g8431	hypothetical protein EMCG_00948	6.38	7.76	8.56	7.46
g8448	lovastatin nonaketide synthase	5.67	6.54	7.86	6.14
g8449	---NA---	5.92	3.93	2.24	1.70
g8477	---NA---	5.56	6.73	7.57	6.80
g8572	benzoate-para-hydroxylase (cytochrome P450)	6.12	2.68	1.60	2.52
g8686	---NA---	7.09	5.90	4.48	5.74
g8700	small oligopeptide OPT family	6.67	4.88	9.30	8.75
g8749	hypothetical protein PENSUB_11874	3.36	4.31	6.45	3.27
g8782	---NA---	7.38	4.19	4.29	4.85
g8787	allophycocyanin subunit beta	7.42	5.88	3.96	4.27
g8788	---NA---	5.34	6.31	8.21	1.30
g8808	conserved hypothetical protein	3.37	3.89	6.56	5.64
g8899	hypothetical protein SBOR_3823	6.37	2.13	3.15	3.40
g8951	---NA---	1.50	-2.41	-6.31	-4.80
g8991	---NA---	7.15	5.85	3.53	4.59
g9445	Oxysterol-binding protein	5.98	1.19	0.74	1.41
g9558	---NA---	4.12	6.19	6.97	6.10
g9639	RNI-like protein	4.91	2.90	9.06	6.14
g9685	hypothetical protein CISG_07311	-2.40	5.19	6.31	6.23
g9738	hypothetical protein PV08_01728	-1.34	5.52	6.42	6.25
g9826	---NA---	5.55	6.27	7.86	7.24
g9834	hypothetical protein M408DRAFT_23609	6.62	5.60	2.94	3.39
g9875	hypothetical protein TEGS_03390	5.80	4.43	0.35	1.04
g9882	amine flavin-containing superfamily	6.06	4.59	0.90	2.44
g9923	hypothetical protein AUEXF2481DRAFT_5316	5.90	3.44	9.62	7.65

g9969	---NA---	-4.47	5.53	5.30	4.92
g10046	---NA---	3.52	3.85	7.22	3.58
g10127	hypothetical protein UMAG_04488	5.57	5.99	7.53	7.43
g10169	hypothetical protein T310_3265	4.46	2.85	7.82	6.54
g10215	oxidoreductase	7.24	5.29	4.57	6.18
g10277	retinol dehydrogenase 12	5.14	2.48	7.45	5.84
g10454	trehalose-6-phosphate hydrolase	2.12	3.92	6.71	2.26
g10485	flavo monooxygenase	1.74	5.31	6.51	3.01
g10514	benzoate 4-monooxygenase cytochrome p450	6.43	5.95	7.58	8.36
g10528	hypothetical protein EPUS_00425	6.25	4.22	6.62	7.99
g10531	hypothetical protein A1O3_04947	4.99	4.84	6.91	7.17
g10556	hypothetical protein W97_02516	2.85	-0.57	-3.25	-3.04
g10672	hypothetical protein A1O3_04947	6.46	6.15	5.79	3.46
g10715	hypothetical protein EPUS_08078	2.75	-2.40	-4.59	-4.06
g10721	catalase	6.55	5.64	8.84	8.33
g10723	cyanate hydratase	6.66	5.06	4.10	3.58
g10815	hypothetical protein PAAG_02813	6.07	6.21	5.52	7.79
g10896	cytochrome P450	4.36	4.02	5.77	6.51
g10897	hypothetical protein ANO14919_040670	5.80	6.18	7.84	7.80
g11025	plasma membrane stress response	6.71	4.88	6.07	1.14
g11090	predicted protein	4.81	6.72	6.68	7.57
g11138	Common central domain of tyrosinase	5.73	5.51	7.11	7.79
g11295	phycocyanin subunit beta	6.82	6.12	8.58	8.45
g11296	phycocyanin subunit alpha	6.88	6.00	5.74	3.60
g11331	O-methyltransferase	3.45	3.99	4.56	7.12
g11430	phosphoribosylformylglycinamide synthase	8.55	10.66	11.57	10.85

g11434	uncharacterized protein PAC_17075	7.40	7.35	6.82	5.03
g11436	related to multidrug resistant	7.30	6.82	6.51	5.39
	related to P450 family sporulation-specific N-formyltyrosine				
g11453	oxidase Dit2	7.81	7.58	7.06	5.51
g11586	hypothetical protein FAVG1_02996	2.04	-4.89	-3.46	-3.99
g11622	hypothetical protein V498_10729	-0.37	-2.83	-5.07	-4.75
g11638	hypothetical protein UCDDS831_g00649	8.57	11.09	11.95	10.78
g11712	methyltransferase	-4.46	3.98	0.62	0.35
g11749	hypothetical protein TSTA_020850	6.33	2.72	4.53	3.13
g11784	carbohydrate-binding module family 13	7.29	5.83	5.45	3.96
g11798	6-hydroxy-D-nicotine oxidase	5.77	4.74	5.23	0.91
g11836	benzoate 4-monooxygenase cytochrome p450	7.82	7.60	8.31	5.10
g11923	hypothetical protein ANO11243_060950	5.65	2.21	3.82	-0.37
g11971	cytochrome P450	-5.41	0.44	-1.60	4.06
g12031	hypothetical protein AUD_4576	5.98	5.79	6.03	7.99
g12157	uncharacterized protein PAC_15196	7.09	6.53	6.37	4.18
g12194	effector candidate	6.18	5.85	5.35	3.09
g12195	Gag-Pol poly	6.96	4.56	4.93	4.60
g12197	gag-pol poly	-2.70	4.13	1.68	6.85
g12199	hypothetical protein WG66_11260	6.51	7.11	6.24	2.46
g12201	hypothetical protein GALMADRAFT_237585	-0.50	-3.69	1.46	2.74
g12202	hypothetical protein SI65_06794	5.45	3.54	6.51	6.30
g12203	gag-pol poly	5.94	0.93	5.64	6.10
g12358	hypothetical protein ASPBRDRAFT_41996	6.47	3.31	6.91	6.93
g12615	---NA---	4.37	6.10	2.54	2.53
g12739	---NA---	8.03	5.40	9.20	8.58

g12743	hypothetical protein T310_3447	3.05	6.24	2.78	2.13
g12744	hypothetical protein HIM_10392	6.23	4.17	7.14	6.77
g12803	hypothetical protein UCRNP2_3765	6.31	2.04	7.19	6.68
g12812	---NA---	3.01	4.87	-0.63	0.61
g12814	Microtubule binding HOOK3	7.22	4.66	8.24	7.86
g12853	Thymidylate kinase	7.66	5.24	8.56	8.25
g12860	predicted protein	-1.55	-6.94	-1.51	-3.91
g12883	kinesin light	6.63	4.23	6.98	6.69
g12915	hypothetical protein A1O1_08442	7.37	5.01	8.38	7.67
g12955	---NA---	5.28	1.67	6.88	6.26
g13010	serine threonine kinase	1.83	0.53	5.38	2.51
g13012	hypothetical protein EPUS_03941	3.55	-0.55	6.11	4.86
g13092	Nonribosomal Peptide Synthase (NRPS)	6.44	4.14	7.18	7.62
g13121	serine threonine- kinase chk1	7.28	5.07	8.82	7.34
g13150	hypothetical protein ASPCAL07495	-2.88	-4.07	1.72	-2.77
g13177	hypothetical protein Z518_00265	6.39	7.68	4.59	5.10
g13209	alpha beta-hydrolase	5.29	6.48	2.89	3.34
g13241	uncharacterized protein CPUR_00784	4.30	5.27	0.97	2.62
g13256	hypothetical protein ASPGLDRAFT_182435	4.91	6.37	3.41	3.36
g13280	---NA---	-4.06	-6.49	0.27	0.60
g13337	Calcium calmodulin-dependent kinase type	4.71	5.95	1.60	2.21
g13476	hypothetical protein PV11_09631	7.61	4.78	7.85	7.65
g13500	hypothetical protein M422DRAFT_246172	6.21	1.97	7.93	7.04
g13508	cell division control 42	4.51	2.27	5.95	5.59
g13644	conserved hypothetical protein	5.79	2.39	6.07	5.74
g13689	---NA---	3.66	0.05	5.77	4.86

g13723	hypothetical protein SISNIDRAFT_484147	6.46	4.19	6.99	6.74
g13749	FAD dependent oxidoreductase	7.27	4.09	8.10	7.49
g13782	photosystem I core	7.75	4.98	7.43	7.50
g13826	ankyrin	-0.44	-6.62	-1.54	-1.13
g13849	hypothetical protein ASPCAL07495	7.48	4.99	8.59	8.64
g13856	MFS general substrate transporter	5.03	-1.50	4.78	4.91
g13857	MFS general substrate transporter	7.44	5.08	7.86	7.49
g13876	Ferritin-like domain	8.03	5.96	10.44	8.82
g13918	---NA---	2.94	-4.46	2.57	3.30
g13931	---NA---	5.70	0.56	5.96	4.40
g14007	hypothetical protein AZE42_07563	6.49	3.09	8.93	7.46
g14008	hypothetical protein ASPGLDRAFT_486713	5.46	-2.12	4.78	0.19
g14053	cytochrome p450	4.66	6.45	3.22	3.57
g14089	---NA---	6.67	2.79	6.67	6.84
g14208	hypothetical protein PV04_05304	6.16	1.98	5.95	6.24
g14238	Sec1 family superfamily	2.28	-3.82	1.68	1.75
g14239	uncharacterized protein CPUR_08621	6.01	1.74	6.63	6.01
g14241	hypothetical protein PSTG_00042	2.91	-2.10	4.88	5.49
g14339	ankyrin repeat	2.54	6.31	3.35	1.94
g14353	---NA---	-2.53	5.10	-0.65	-3.28
g14355	---NA---	7.79	5.14	9.34	8.36
g14381	hypothetical protein TBLA_0E01410	4.19	-3.53	1.12	-1.47
g14404	hypothetical protein VE00_05995	5.92	1.25	5.90	5.59
g14409	kinesin light chain	6.27	7.00	4.04	3.59
g14423	uncharacterized protein PAC_01090	7.14	4.31	7.08	7.59
g14470	hypothetical protein A4X13_g7656	6.31	2.00	6.12	4.35

g14483	hypothetical protein GYMLUDRAFT_247080	2.14	5.95	2.88	3.00
g14501	---NA---	-2.80	-3.99	0.80	-1.09
g14525	P30 adhesin	7.09	5.25	7.64	7.73
g14543	reverse transcriptase	8.42	6.37	10.41	9.52
g14662	---NA---	2.32	1.29	5.05	5.60
g14684	---NA---	6.36	3.81	6.83	7.04
g14706	---NA---	5.14	7.11	5.39	4.14
g14707	NAD(P)-binding protein	2.28	-2.12	4.03	4.58
g14708	hypothetical protein CC84DRAFT_1195088	4.21	-1.18	2.09	5.20
g14709	hypothetical protein H101_03049	7.03	4.25	5.31	7.54
g14729	reducing type I polyketide synthase	-0.76	-3.90	0.65	2.00
g14731	hypothetical protein BofuT4_P088950.1	3.07	6.26	2.70	0.76
g14736	hypothetical protein Hypma_11659	6.46	7.72	6.18	4.31
g14762	hypothetical protein K441DRAFT_650523	3.79	-1.79	2.76	5.30
g14802	hypothetical protein AOQ84DRAFT_191815	4.14	-2.91	2.25	5.69
g14807	hypothetical protein K457DRAFT_39792, partial	7.07	3.64	6.33	7.26
g14808	---NA---	5.81	1.96	5.48	6.51
g14809	hypothetical protein M378DRAFT_858718	0.74	-4.35	-0.09	2.39
g14891	Retrovirus-related Pol poly from transposon TNT 1-94	5.24	7.06	4.92	3.68
g14893	---NA---	5.70	5.67	5.66	0.15
g14896	hypothetical protein SS1G_06147	3.39	-0.20	4.94	5.70
g14910	hypothetical protein BAUCODRAFT_145551	4.27	6.55	6.21	1.57
g14944	---NA---	4.92	5.67	3.87	-1.81
g14958	---NA---	-1.08	5.37	0.65	-1.46
g14993	ankyrin	5.87	7.28	5.95	4.27
g14998	---NA---	3.09	-1.53	4.12	4.73

g15000	hypothetical protein ST47_g8211	4.56	2.11	6.03	6.75
g15007	---NA---	1.94	-2.15	3.48	4.28
g15043	GMC oxidoreductase,	4.84	5.85	1.65	-1.16
g15061	hypothetical protein AGAB12DRAFT_176739	5.27	3.66	6.52	4.18
g15072	hypothetical protein PTTG_27154	2.16	3.41	2.64	5.42
g15122	WD40 repeat	5.64	6.53	6.14	4.42
g15136	von Willebrand factor, type A	5.47	6.16	6.55	4.76
g15153	benzoate 4-monooxygenase cytochrome P450	7.11	7.01	7.70	6.10
g15177	hypothetical protein ALT_3415	4.08	4.45	2.91	5.37
g15178	hypothetical protein PV05_09409	3.08	1.88	0.18	3.55
g15180	atg7p	0.03	-3.16	-1.94	1.11
g15181	MAPKKK cascade kinase regulator STE50	4.44	3.79	5.87	3.53
g15202	reducing type I polyketide synthase	1.40	0.98	-2.14	0.80
g15272	hypothetical protein ASPVEDRAFT_41603, partial	1.55	-3.12	-4.07	-0.61
g15282	---NA---	5.54	7.07	6.03	4.11
g15289	hypothetical protein Z518_01135	1.50	3.29	0.45	4.20
g15292	hypothetical protein ACJ73_02205	6.74	7.82	6.62	4.96
	Saccharomyces cerevisiae YJR138W IML1 Subunit of the SEA				
g15293	(Seh1-associated) complex	5.77	4.59	5.18	2.14
g15303	hypothetical protein PV08_06397	3.43	4.56	5.47	1.16
g15304	---NA---	4.16	5.54	4.80	6.37
g15305	hypothetical protein DOTSEDRAFT_37712	1.86	4.52	5.40	2.90
g15321	Tetratricopeptide-like helical	5.72	7.50	6.56	4.89
g15357	WD40 repeat	6.21	6.14	5.60	7.12
g15404	---NA---	4.42	6.32	6.70	4.86
g15422	Cytochrome P450	6.75	5.47	6.24	8.06

g15423	cytochrome P450	5.58	5.01	4.39	6.51
g15467	reducing type I polyketide synthase	5.18	4.60	6.19	4.26
g15478	hypothetical protein HMPREF1541_07223	-2.47	-3.66	-2.95	1.99
g15501	hypothetical protein PFICI_15058	-3.35	-0.32	0.16	-3.05
g15507	---NA---	8.11	6.63	6.13	8.14
g15539	Clavaminic synthase	5.41	3.41	2.00	5.43
g15557	hypothetical protein TOPH_06261	-1.62	0.81	-2.88	0.21
g15587	uncharacterized protein PAC_19084	0.31	-0.11	-4.30	2.68
g15595	hypothetical protein L228DRAFT_270417	0.53	1.73	-3.03	1.44
g15614	GroES-like protein	-0.30	2.03	0.76	-1.94
g15647	hypothetical protein K432DRAFT_430842	2.76	-0.02	0.70	4.35
g15656	benzoate 4-monooxygenase cytochrome P450	5.63	3.77	1.49	5.47
g15670	hydrolase, partial	1.52	-0.65	-0.80	4.11
g15713	Serine/threonine/dual specificity protein kinase, catalytic domain	2.85	3.95	0.04	5.56
g15744	hypothetical protein BC1G_01702	4.97	4.47	4.71	1.63
g15795	hypothetical protein I317_07682	-0.05	5.13	4.09	0.54
g15825	hypothetical protein A4X13_g8847, partial	2.57	-2.57	-2.93	0.50
g15875	reducing type I polyketide synthase	2.42	3.20	2.60	5.52
g15876	CIC11C00000004402 [	4.79	5.48	5.09	1.68
g15898	hypothetical protein COCSADRAFT_48148, partial	-3.41	0.84	-1.20	3.47
g15919	hypothetical protein PIIN_04925	-1.00	3.25	-0.05	5.15
g15926	hypothetical protein TREMEDRAFT_66424	2.73	0.08	-3.89	-0.01
g15951	related to monooxygenase	4.73	6.92	7.05	5.19
g15975	alpha beta-Hydrolase	6.61	6.72	6.40	2.26
g16008	---NA---	6.09	5.75	6.10	3.70
g16010	hypothetical protein EPUS_00568	4.51	3.22	6.06	3.83

g16027	hypothetical protein TREMEDRAFT_62576	1.11	2.67	5.71	0.62
g16118	---NA---	-3.38	-4.57	0.91	-1.76
g16163	uncharacterized protein VP01_908g6	3.62	4.47	4.79	0.65
g16164	---NA---	2.80	2.91	5.65	1.56
g16168	hypothetical protein L228DRAFT_242344	3.36	-2.15	-1.43	3.89
g16179	---NA---	4.24	-1.84	-1.12	3.23
g16181	---NA---	4.03	-3.90	-3.11	1.17
g16184	---NA---	-3.44	1.73	-0.19	4.56
g16274	---NA---	6.15	6.36	7.55	5.57
g16282	mfs multidrug transporter	4.62	5.56	4.45	0.34
g16328	---NA---	6.16	5.98	5.42	2.73
g16343	DNA helicase recq5,	-2.74	0.74	-0.15	4.06
g16354	hypothetical protein V502_09950	3.58	2.51	3.22	6.28
g16358	hypothetical protein LY89DRAFT_739923	0.38	-1.15	2.95	-3.05
g16365	related to ribokinase	4.08	3.86	4.52	0.97
g16395	---NA---	-1.69	1.98	4.92	0.34

Table S26: List of 76 DEGs between R (au-10) vs. R (su-12).

Gene id.	Description	R (au-2010) (log2 RPM)	R (su-2012) (log2 RPM)
g365	---NA--- hypothetical protein	5.8	7.6
g425	AZE42_02052	4.9	0.5
g1316	-ribulosamine 3- chloroplastic	6.5	-2.5
g2305	---NA--- hypothetical protein	4.9	6.9
g4526	K441DRAFT_186449	5.4	7.0
g4676	conidiation-specific 6 hypothetical protein	4.6	7.0
g5122	HMPREF1120_05181	4.8	-0.8
g5168	Subtilisin-like protease	6.1	0.0
g5542	---NA---	5.2	-1.1
g5546	---NA---	5.7	7.7
g6342	---NA---	4.8	6.9
g6620	---NA--- hypothetical protein	3.8	6.0
g7881	PV08_05142 hypothetical protein	5.7	7.6
g8210	SS1G_04281	5.2	7.0
g8211	---NA--- outer membrane	2.8	5.8
g8218	autotransporter barrel hypothetical protein	4.3	6.5
g8431	EMCG_00948	6.7	4.5
g9550	---NA--- hypothetical protein	5.7	7.5
g9738	PV08_01728	6.9	4.9
g10189	---NA---	-0.9	-6.1
g10277	retinol dehydrogenase 12 trehalose-6-phosphate	5.5	8.3
g10454	hydrolase MFS general substrate	6.3	2.1
g10745	transporter hypothetical protein	4.9	6.8
g11148	V501_04538	4.5	-1.1
g11434	uncharacterized protein	1.7	5.7

	PAC_17075		
g11711	HC-toxin efflux carrier TOXA	4.5	7.5
g11712	methyltransferase	4.2	7.7
	hypothetical protein		
g11713	M437DRAFT_59426	4.4	7.5
	carbohydrate-binding module		
g11784	family 13	6.5	-2.8
	hypothetical protein		
g12023	CPC735_009070	5.3	1.1
g12243	---NA---	1.6	-4.2
g12362	---NA---	6.5	4.5
g12494	exo-beta- -glucanase	6.8	4.0
	Cytochrome P450 pisatin		
g12696	demethylase	2.2	-4.9
g12860	predicted protein	5.7	-0.4
g13137	Cytochrome P450	5.8	3.3
g13138	glutathione S-	5.9	2.0
	hypothetical protein		
g13256	ASPGLDRAFT_182435	4.9	7.8
g13257	catalytic	5.4	7.7
g13505	---NA---	0.8	-6.4
g13749	FAD dependent oxidoreductase	4.4	7.8
g13757	TPR domain-containing	5.2	-5.0
	uncharacterized protein		
g13758	CPUR_00784	5.0	-3.0
	hypothetical protein		
g13901	VE00_10792	7.0	5.1
g14186	TPR domain-containing	4.0	-4.8
g14562	---NA---	1.6	-4.3
	hypothetical protein		
g14609	EPUS_00365	6.7	4.9
g14706	---NA---	5.1	-0.4
g14707	NAD(P)-binding protein	5.7	-2.8
	hypothetical protein		
g14708	CC84DRAFT_1195088	5.5	-1.2
	hypothetical protein		
g14728	CC77DRAFT_1057376	1.8	-4.1
g14808	---NA---	5.1	-2.9
	hypothetical protein		
g14809	M378DRAFT_858718	5.7	-4.6
g14869	---NA---	6.6	4.5

g14993	ankyrin	5.6	-4.1
g14998	---NA---	7.0	-0.2
g15008	---NA---	1.6	-5.1
g15122	WD40 repeat hypothetical protein	2.8	-5.2
g15292	ACJ73_02205	5.1	-0.7
g15390	---NA---	4.7	7.0
	hypothetical protein		
g15420	VE00_10790	6.5	4.6
g15422	Cytochrome P450	5.3	-1.8
g15423	cytochrome P450 hypothetical protein	5.5	-2.0
g15451	V494_00628	5.9	7.7
g15502	lectin-2 hypothetical protein	6.3	1.7
g15608	ASPWEDRAFT_172669	4.9	-1.2
g15614	GroES-like protein benzoate 4-monooxygenase	4.8	7.3
g15656	cytochrome P450 Serine/threonine/dual specificity protein kinase,	5.3	-3.0
g15713	catalytic domain hypothetical protein	5.3	-3.1
g15831	VE03_06097 hypothetical protein	4.5	-1.3
g16168	L228DRAFT_242344	5.5	-4.3
g16179	---NA---	5.4	1.1
g16184	---NA---	6.0	-1.3
g16343	DNA helicase recq5,	5.1	-0.8
g16365	related to ribokinase	5.9	-2.1
g16377	---NA---	4.9	-2.2

Table S27: List of 34 DEGs between TB (su-12) vs. TW-G (su-12), and the 7-background variation DEGs are highlighted in red.

Gene id.	Description	TB-2012 (log2 RPM)	TW (G)-2012 (log2 RPM)
g530	conidial pigment biosynthesis oxidase Abr1 brown 1	6.10	6.92
g2304	---NA---	5.77	6.84
g2698	A5178_ARTBC ame: Full=Uncharacterized secreted ARB_05178 Flags: Precursor	3.68	5.00
g3540	hypothetical protein GYMLUDRAFT_706765	5.35	6.34
g3893	hypothetical protein GLAREA_08855	6.53	7.50
g4354	---NA---	4.58	5.98
g4509	1,3,6,8-tetra-HN reductase	5.63	7.17
g4510	transcription factor	5.89	6.52
g4607	hypothetical protein OIADMADRAFT_35369	4.25	6.10
g5239	BUB1-like protein	1.93	4.44
g6533	---NA---	6.51	6.02
g7466	chitin biosynthesis	4.71	5.44
g9715	---NA---	2.38	4.84
g9822	ATP-dependent RNA helicase ded1,	-4.35	1.10
g10296	hypothetical protein AOQ84DRAFT_351571	3.90	5.09
g10967	---NA---	5.76	7.02
g11138	Common central domain of tyrosinase	5.70	6.64
g11259	MFS multidrug resistance transporter,	5.31	5.87
g11434	uncharacterized protein PAC_17075	-2.60	2.14
g11435	uncharacterized protein PAC_10069	0.54	3.47
g11711	HC-toxin efflux carrier TOXA	4.44	5.56
g11712	methyltransferase	2.02	5.63
g11713	hypothetical protein M437DRAFT_59426	2.25	5.18
g12991	uncharacterized protein FFUJ_00300	1.26	3.41
g13256	hypothetical protein ASPGLDRAFT_182435	3.67	5.84
g13257	catalytic	4.25	5.68

g13486	hypothetical protein PV07_12346	3.10	5.02
g13487	beta-ig-h3 fasciclin	5.08	6.90
g13734	hypothetical protein K432DRAFT_290587	2.32	4.11
g13788	hypothetical protein VF21_00684	0.14	3.16
g13789	hypothetical protein AOQ84DRAFT_329710	1.21	3.32
g15072	hypothetical protein PTTG_27154	4.51	6.45
g15680	hypothetical protein ASPWEDRAFT_28951	-1.21	2.75
g16121	hypothetical protein, partial	3.16	4.88

Table S28: List of dioxygenases and lipases genes in *P. membranacea*.

Gene id.	Description	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
g93	autophagy related lipase	5.61	4.67	4.78
g473	lipase class 3	5.49	6.22	6.41
g518	small subunit of phenylpropionate dioxygenase protein	4.79	6.50	7.68
g667	3-hydroxyanthranilate –dioxygenase	4.41	5.74	5.93
g807	triacylglycerol lipase	4.60	4.78	4.85
g1009	triacylglycerol lipase	5.47	6.46	6.11
g1046	dihydroxy–3–keto–5–methylthiopentene dioxygenase	7.05	6.95	6.36
g2429	cytosolic phospholipase a2 zeta protein	4.85	4.48	3.98
g2781	phospholipase d active site motif protein	7.10	6.28	6.30
g3232	patatin–like phospholipase	4.76	4.62	4.54
g3233	patatin phospholipase a2–related protein	5.34	4.67	4.88
g4278	2–nitropropane dioxygenase protein	6.67	5.85	6.07
g4344	lipase gdsI	5.02	5.91	5.73
g5160	patatin phospholipase a2–related protein	4.92	5.67	5.73
g5586	lysophospholipase	4.78	3.30	3.93
g5622	trimethyllysine dioxygenase	3.85	4.23	4.29
g6187	fe ii 2–oxoglutarate–dependent dioxygenase protein	3.71	0.39	-2.99
g6224	lipase	4.18	5.06	4.59
g6799	lipase esterase family protein	5.93	5.25	4.95
g6966	indoleamine –dioxygenase subfamily	5.08	5.55	5.11
g8011	cysteine dioxygenase type i family protein	4.42	5.41	3.20

g8129	extracellular lipase	6.90	6.30	6.66
g8680	2-nitropropane dioxygenase	4.71	4.49	3.85
g10458	lysophospholipase-like protein	5.65	4.16	4.05
g10490	gamma-butyrobetaine dioxygenase	2.72	2.14	-0.54
g10501	phytanoyl- dioxygenase	4.94	5.58	6.08
g10516	secretory lipase protein	4.82	6.32	4.94
g11493	carotenoid (Selker)-cleavage dioxygenase	6.50	6.42	6.35
g11564	phospholipase c protein	6.14	6.22	6.45
g11888	lipase atg15	6.09	5.80	5.89
g12459	thymine dioxygenase	4.41	5.61	5.77
g13824	alpha-ketoglutarate-dependent taurine dioxygenase	3.63	2.09	3.19
g14161	phytanoyl- dioxygenase family protein	3.24	2.39	3.49
g15031	lysosomal acid lipase cholesteryl ester hydrolase	6.36	6.39	6.07
g15539	taurine catabolism dioxygenase protein	5.47	5.91	3.79
g15942	phytanoyl- dioxygenase family protein	0.03	5.36	4.51

Appendix I: Southern blotting protocol.

1. Digest DNA with appropriate restriction enzyme overnight.
2. Run digested DNA on a 1–1.5% agarose gel at very low voltage overnight in a cold room.
3. Next day briefly check the progress under UV. Take a picture of the gel with a ruler on one of the lateral sides. Proceed to blotting.
4. (optional) Trim extra gel off. Measure and write down the size of the final gel before the blotting procedure.
5. Gently shake the gel in 0.25 M HCl (around 10 min, until bromophenol blue turns yellow).
6. Rinse in H₂O.
7. Gently shake in 0.4N NaOH for around 30 min.
8. Proceed to alkaline transfer using 0.4 M NaOH according to the standard procedure.
9. Blot in 0.4 M NaOH for 2–16 hours until the gel will be completely shrunk.
10. Draw with a pencil the borders of the gel on the membrane.
11. Label the membrane.
12. Rinse in 2×SSC for 5 min.
13. Trim all extra membrane.
14. Wrap in Saranwrap and store desiccated at NT or at –20°C.
15. Label your probe using random primed labelling with Klenow.
16. Denature your linear DNA by heating for 10 min at 95°C and immediately put on ice.
17. Assemble the labeling reaction as follows.

Linear template DNA	10 ng – 2 µg DNA
Nucleotides, final concentration	25 µM of dATP, dCTP, dTTP each
Labeled nucleotides, final concentration	$\alpha(^{32}\text{P})$ dGTP 3,000 Ci/mmol), 50 µCi (1.85 MBq)
10× hexanucleotides mix (62.5 A260 U/ml)	2 µl
Klenow enzyme	2 U
10× Labeling buffer	2 µl
H ₂ O	add up to 20 µl
Incubation	30 min at 37°C

Labelling buffer (10×): 500 mM Tris (pH 7.2), 100 mM MgCl₂, 1 mM DTT, and 2 mg/ml BSA.

18. Terminate the reaction by adding 2 µl 0.2 M EDTA (pH 8.0) and/or heating to 65°C for 10 minutes.

19. Remove the un-incorporated ^{32}P with sephadex G-50 spin columns. After adding the reaction mix to the column, wash the tube with 100 μl of STE to the column.

20. STE (1 $\times$, store at 4°C)

	Final	Stock	1ml	5ml	10ml	50ml
NaCl	0.1M	5M	20 μl	100 μl	200 μl	1.0ml
Tris Cl, pH8.0	10mM	1M	10 μl	50 μl	100 μl	0.5ml
EDTA, pH8.0	1mM	0.5M	2 μl	10 μl	20 μl	100 μl
H ₂ O mQ			0.968ml	4.84ml	9.68ml	48.4ml

21. Assay the reaction by using a hand-held Geiger counter. A good reaction registers between 1 and 2 kcps (thousand counts per sec) at a distance of about an inch from the phototube. If the reaction does not work, reading would be low, possibly 10-fold decrease, indicating the probe would not be useful.

22. Proceed to hybridization according to Hybond-XL protocol.

Appendix J: Analyses of repeat elements, RIP indices and *rid* gene identification in *P. membranacea* genome.

Table S29: Repeat elements in *P. membranacea* genome.

=====			
file name:	454AllContigs.fna		
sequences:	150581		
total length:	135511117 bp	(135511117 bp excl N/X-runs)	
GC level:	44.89 %		
bases masked:	5381800 bp (3.97 %)		
=====			
	number of elements*	length occupied	percentage of sequence

Retroelements	12811	3895219 bp	2.87 %
SINEs:	0	0 bp	0.00 %
Penelope	0	0 bp	0.00 %
LINEs:	88	9984 bp	0.01 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	0	0 bp	0.00 %
R1/LOA/Jockey	10	493 bp	0.00 %
R2/R4/NeSL	0	0 bp	0.00 %
RTE/Bov-B	0	0 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	12723	3885235 bp	2.87 %
BEL/Pao	0	0 bp	0.00 %
Ty1/Copia	779	199211 bp	0.15 %
Gypsy/DIRS1	11944	3686024 bp	2.72 %
Retroviral	0	0 bp	0.00 %
DNA transposons	93	10680 bp	0.01 %
hobo-Activator	3	600 bp	0.00 %
Tc1-IS630-Pogo	69	7828 bp	0.01 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	17	1699 bp	0.00 %
PiggyBac	0	0 bp	0.00 %
Tourist/Harbinger	1	157 bp	0.00 %
Other (Mirage, P-element, Transib)	0	0 bp	0.00 %
Rolling-circles	0	0 bp	0.00 %
Unclassified:	45	9242 bp	0.01 %
Total interspersed repeats:		3915141 bp	2.89 %
Small RNA:	161	30289 bp	0.02 %
Satellites:	0	0 bp	0.00 %
Simple repeats:	7684	385647 bp	0.28 %
Low complexity:	18556	1052112 bp	0.78 %
=====			

```

=====
file name: Pmem v1.1 (rhiz_repscaff_fh.fasta)
sequences:          3033
total length:    43152721 bp (43042200 bp excl N/X-runs)
GC level:        44.38 %
bases masked:    1050958 bp ( 2.44 %)
=====

```

	number of elements*	length occupied	percentage of sequence

SINEs:	0	0 bp	0.00 %
ALUs	0	0 bp	0.00 %
MIRs	0	0 bp	0.00 %
LINEs:	0	0 bp	0.00 %
LINE1	0	0 bp	0.00 %
LINE2	0	0 bp	0.00 %
L3/CR1	0	0 bp	0.00 %
LTR elements:	0	0 bp	0.00 %
ERV1	0	0 bp	0.00 %
ERV1-MaLRs	0	0 bp	0.00 %
ERV_classI	0	0 bp	0.00 %
ERV_classII	0	0 bp	0.00 %
DNA elements:	0	0 bp	0.00 %
hAT-Charlie	0	0 bp	0.00 %
TcMar-Tigger	0	0 bp	0.00 %
Unclassified:	3152	542083 bp	1.26 %
Total interspersed repeats:		542083 bp	1.26 %
Small RNA:	0	0 bp	0.00 %
Satellites:	0	0 bp	0.00 %
Simple repeats:	16027	634021 bp	1.47 %
Low complexity:	0	0 bp	0.00 %
=====			

Table S30: RIP analysis.

Sequences		No of bases & No of sequences	ApA	ApC	ApG	ApT	CpA	CpC	CpG	CpT	GpA	GpC	GpG	GpT	TpA	TpC	TpG	TpT	TpA/ApT (>0.89)	(CpA+TpG)/ (ApC+GpT) < 1.03
Genome	P.mem454Allcontigs	150,581 seqs	0.07	0.05	0.07	0.09	0.07	0.04	0.05	0.07	0.07	0.06	0.05	0.05	0.07	0.07	0.07	0.07	0.85	1.23
	P.mem v1.0	14,674 seqs	0.07	0.05	0.07	0.08	0.07	0.05	0.05	0.07	0.07	0.06	0.05	0.06	0.06	0.07	0.07	0.07	0.79	1.25
	P.mem v1.1	3,033 seqs	0.07	0.05	0.07	0.08	0.07	0.05	0.04	0.07	0.07	0.05	0.05	0.05	0.07	0.07	0.07	0.07	0.81	1.24
	DNA	103bp x 3	0.10	0.08	0.04	0.16	0.05	0.03	0.02	0.06	0.05	0.02	0.02	0.03	0.19	0.03	0.03	0.06	1.15	0.73
	DNA-Ginger	202bp x 4	0.06	0.07	0.07	0.10	0.10	0.03	0.03	0.05	0.08	0.05	0.04	0.05	0.06	0.07	0.08	0.04	0.63	1.37
DNA transposons	DNA-hAT	225bp x 2	0.08	0.06	0.08	0.08	0.08	0.02	0.03	0.07	0.10	0.06	0.04	0.04	0.04	0.06	0.09	0.06	0.57	1.66
	DNA-hAT-Restless	126bp x 1	0.08	0.06	0.08	0.08	0.08	0.02	0.03	0.07	0.10	0.06	0.04	0.04	0.04	0.06	0.09	0.06	0.57	1.66
	DNA-MuDR	194bp x 12	0.06	0.07	0.05	0.10	0.10	0.03	0.03	0.09	0.05	0.07	0.03	0.02	0.07	0.08	0.06	0.08	0.74	1.67
	DNA-TcMar-Fot1	520bp x 22	0.06	0.08	0.05	0.11	0.09	0.03	0.03	0.06	0.07	0.04	0.05	0.04	0.08	0.06	0.07	0.08	0.74	1.32
	DNA-TcMar-Pogo	47bp x 1	0.06	0.08	0.05	0.11	0.09	0.03	0.03	0.06	0.07	0.04	0.05	0.04	0.08	0.06	0.07	0.08	0.74	1.32
	DNA-TcMar-Tc1	88bp x 2	0.10	0.04	0.08	0.11	0.04	0.05	0.04	0.04	0.11	0.04	0.09	0.04	0.07	0.04	0.06	0.06	0.67	1.38
	16_DNA_TcMar-Mariner	75bp x 10	0.08	0.03	0.02	0.11	0.02	0.05	0.03	0.09	0.12	0.06	0.08	0.02	0.03	0.05	0.14	0.07	0.31	3.17
	LINE-1	68bp x 6	0.08	0.03	0.11	0.09	0.07	0.06	0.03	0.05	0.11	0.08	0.08	0.02	0.06	0.03	0.06	0.03	0.67	2.67
Retroelements	LINE Penelope	62bp x 2	0.22	0.19	0.00	0.00	0.01	0.25	0.00	0.16	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.07
	LINE-Tad1	122bp x 24	0.06	0.06	0.09	0.07	0.09	0.06	0.03	0.06	0.08	0.06	0.06	0.04	0.04	0.07	0.07	0.05	0.58	1.59
LTR Elements	LTR-Copia	2205 x 418	0.07	0.07	0.07	0.08	0.08	0.04	0.04	0.07	0.08	0.06	0.05	0.05	0.06	0.07	0.07	0.06	0.73	1.29
	LTR-Gypsy	1924 x 2000	0.08	0.07	0.07	0.09	0.08	0.05	0.03	0.07	0.07	0.05	0.05	0.04	0.07	0.06	0.06	0.07	0.80	1.29
	RC-Helitron	422 x 9	0.07	0.07	0.05	0.10	0.08	0.03	0.03	0.07	0.05	0.03	0.04	0.07	0.08	0.07	0.07	0.09	0.86	1.07

Table S31: *rid* (RIP defective) homolog in *P. membranacea* and other fungi.

Organism	Class	Order	RID
<i>Arthonia cinnabarina</i>	Arthoniomycetes	Arthoniales	Present
<i>Leptosphaeria maculans</i>	Dothideomycetes	Pleosporales	Present
<i>Cochliobolus carbonun</i>	Dothideomycetes	Pleosporales	Present
<i>Uncinocarpus reesei</i>	Eurotiomycetes	Hypocreales	Present
<i>Endocarpon pallidulum</i>	Eurotiomycetes	Verrucariales	Present
<i>Aspergillus fumigatus</i>	Eurotiomycetes	Eurotiales	DmtAf (rid homolog)
<i>Aspergillus oryzae</i>	Eurotiomycetes	Eurotiales	Present
<i>Aspergillus nidulans</i>	Eurotiomycetes	Eurotiales	DmtA (rid homolog)
<i>Ascobolus immersus</i>	Eurotiomycetes	Pezizales	Masc1 (rid homolog)
<i>Cladonia grayi</i>	Lecanoromycetes	Lecanorales	Present
<i>Peltigera membranacea</i>	Lecanoromycetes	Peltigerales	Absent
<i>Lobaria pulmonaria</i>	Lecanoromycetes	Peltigerales	Present
<i>Acarospora strigosa</i>	Lecanoromycetes	Acarosporales	Present
<i>Dibaeis baeomyces</i>	Lecanoromycetes	Lecanorales	Present
<i>Graphis scripta</i>	Lecanoromycetes	Ostropales	Present
<i>Physcia stellaris</i>	Lecanoromycetes	Lecanorales	Present
<i>Xanthoria parietina</i>	Lecanoromycetes	Teloschistales	Present
<i>Usnea florida</i>	Lecanoromycetes	Lecanorales	Present
<i>Botrytis fuckeliana</i>	Leotiomycetes	Helotiales	Present
<i>Peltula cylindrica</i>	Lichinomycetes	Lichinales	Present
<i>Neurospora crassa</i>	Sordariomycetes	Sordariales	Present
<i>Neurospora tetrasperma</i>	Sordariomycetes	Sordariales	Present
<i>Neurospora intermedia</i>	Sordariomycetes	Sordariales	Present
<i>Sordaria brevicollis</i>	Sordariomycetes	Sordariales	Present
<i>Magnaporthe grisea</i>	Sordariomycetes	Magnaporthales	Present
<i>Gibberella zeae (Fusarium graminearum)</i>	Sordariomycetes	Hypocreales	Present
<i>Trichoderma reesei (Hypocrea jecorina)</i>	Sordariomycetes	Hypocreales	Present
<i>Verticillium albo-atrum</i>	Sordariomycetes	Hypocreales	Present

Appendix K: BS-Seq analysis.

Table S32: Summary of BS-Seq sequencing.

Samples	Libraries	Reads	Total reads analysed	Mapped reads	Mapped reads %	Not Mapped reads	Not Mapped reads %	Average length	Average coverage
Rhizines (Replicate 1)	R 18.1	6,538,712	27,214,886	4,654,196	17	22,560,690	83	96	10.37
	R 18.2	6,538,712							
Rhizines (Replicate 2)	R 4.1	7,068,731							
	R 4.2	7,068,731							
Rhizines (Control)	R.C.1	2,435,706	4871412	101,188	2	4,770,224	97.92	94	0.22
	R.C.2	2,435,706							
Thallus Body (Replicate 1)	TB 4.1	7,979,667	26,600,702	9,068,199	34	17,532,503	66	101	21.16
	TB 4.2	7,979,667							
Thallus Body (Replicate 2)	TB 6.1	5,320,684	559,090	210,690	4	5,348,400	96.21	96	0.46
	TB 6.2	5,320,684							
Thallus Body (Control)	TB C.1	279,545							
	TB C.2	279,545							
Thallus Whole (Replicate 1)	TW 18.1	6,058,937	31,543,593	9,309,689	30	22,233,883	70	101	21.71
	TW 18.2	6,058,937							
Thallus Whole (Replicate 2)	TW 8.1	9,712,849	5,278,274	174,209	3	5,104,065	65	96.7	0.38
	TW 8.2	9,712,849							
Thallus Whole (Control)	TW C.1	2,639,137							
	TW C.2	2,639,137							

Table S33: Overview of BS reads mapped.

	Rhizines		Rhizines control		Thallus Body		Thallus Body control		Thallus Whole		Thallus Whole control	
	Single Reads	Read Pairs	Single Reads	Read Pairs	Single Reads	Read Pairs	Single Reads	Read Pairs	Single Reads	Read Pairs	Single Reads	Read Pairs
Total:	4,654,196	1,995,412	101,188	0	9,068,199	3,684,532	210,690	3	9,309,689	3,768,847	174,209	0
Duplicate:	2,825,112	1,240,841	2500	0	1,846,684	671,114	7,382	0	1,160,427	357,853	5,017	0
Non-specific:	0	0	0	0	0	0	0	0	0	0	0	0
From broken pair:	0		0		0		0		0		0	
Included in analysis:	1,829,084	754,571	98,688	0	7,221,515	3,013,418	203,308	3	8,149,262	3,410,994	169,192	0
Mapped to CT-converted reference:	884,808	379,257	48,861	0	3,524,626	1,521,990	101,125	2	3,978,285	1,721,367	83,927	0
- as CT-converted read:	448,055		24,395		1,748,933		50,861		1,987,969		42,226	
- as GA-converted read:	436,753		24,466		1,775,693		50,264		1,990,316		41,701	
Mapped to GA-converted reference:	944,276	375,314	49,827	0	3,696,889	1,491,428	102,183	1	4,170,977	1,689,627	85,265	0
- as CT-converted read:	473,928		24,812		1,831,608		51,244		2,081,436		42,633	
- as GA-converted read:	470,348		25,015		1,865,281		50,939		2,089,541		42,632	

The table gives an overview of sequences analysed. The column 'Single Reads' counts individual reads, i.e. single-end reads once and paired-end reads twice. The column 'Read Pairs' counts each read pair once. **Total:** The total number of reads/pairs in the input data set. **Duplicate:** The number of duplicate reads/pairs. A single read is called a duplicate when its mapping coordinates are identical to those of another single read. A read pair is called duplicate, when the mapping coordinates of the fragment (the outer mapping coordinates of both reads) are identical to another read pair. **Non-specific:** Reads/pairs that had more than one optimal mapping. **From broken pair:** Single reads that were mapped as single reads but originated from a read pair. **Included in analysis:** How many reads/pairs were included in the methylation call analysis after filtering of aforementioned reads/pairs. **Mapped to CT-converted reference:** The number of reads/pairs that were mapped to the CT-conversion of the reference. **Mapped to GA-converted reference:** The number of reads/pairs that were mapped to the GA-conversion of the reference. (Reproduced from BS-Seq Methylation caller report generated using CLC Genomics Workbench v 9.0.1 software).

Table S34: DNA cytosine methylation levels in rhizines (R), thallus body (TB) and thallus whole (TW).

	Tissues	Methylation contexts	Analysed instances	Unmethylated	Methylated	% of mCs
Whole Genome	R	CG	3,102,352	3,663,804	402,666	9.9
		CHG	2,844,710	3,322,325	599,749	15.29
		CHH	10,622,781	12,341,531	2,713,615	18.02
	TB	CG	3,515,014	18,769,197	1,773,275	8.63
		CHG	3,227,878	16,945,137	2,538,437	13.03
		CHH	12,080,820	57,455,980	10,948,659	16.01
	TW	CG	3,514,864	19,686,782	2,060,914	9.48
		CHG	3,227,568	17,645,148	3,047,468	14.73
		CHH	12,075,750	58,810,081	13,345,109	18.5
Exon	R	CG	620,011	623,106	16,881	2.64
		CHG	535,294	531,713	14,880	2.72
		CHH	1,756,703	1,656,420	48,742	2.86
	TB	CG	701,885	3,380,436	90,571	2.61
		CHG	608,040	2,883,551	79,162	2.67
		CHH	1,996,813	8,495,767	253,498	2.9
	TW	CG	701,876	3,587,364	66,210	1.81
		CHG	607,999	3,040,320	58,986	1.9
		CHH	1,996,871	8,749,650	185,620	2.08
Intron	R	CG	237,865	278,072	19,665	6.6
		CHG	253,470	286,850	26,764	8.53
		CHH	950,222	1,026,047	100,702	8.94
	TB	CG	268,850	1,381,878	89,442	6.08
		CHG	287,668	1,431,066	123,339	7.93
		CHH	1,082,353	4,738,294	438,648	8.47
	TW	CpG	268,986	1,436,763	104,217	6.76
		CHG	287,793	1,481,365	146,115	8.98
		CHH	1,082,624	4,755,137	524,728	9.94

The table gives an overview of methylation level of each cytosine in the tissues. The column 'analysed reference positions' refers to the number of cytosines in the reference sequence that were included in the analysis with respect to the methylation context. 'Unmethylated' is the number of reads that overlapped the cytosines included in the analysis and indicated no methylation. 'Methylated' refers to the number of reads that overlapped the cytosines included in the analysis and indicated methylation. (Adapted from BS-Seq Methylation caller report generated using CLC Genomics Workbench 9.0 software). The bisulfite treatment converts unmethylated cytosines (Cs) to thymines (Ts) after PCR amplification, the methylation level at each cytosine was estimated as $\#C/(\#C + \#T)$, where $\#C$ is the number of methylated reads and $\#T$ is the number of unmethylated reads (Montanini et al. 2014) .

Table S35: List of genes and the methylated sequence contexts of differentially methylated regions for exons.

S.No	Gene id	CG	CHG	CHH	Total
1	g10258			1	1
2	g10601	1	1	1	3
3	g1061		1		1
4	g10923			1	1
5	g10979	1		1	2
6	g11068		1		1
7	g1178	1		1	2
8	g11987			1	1
9	g12344	1	1	1	3
10	g12493			1	1
11	g12618			1	1
12	g12634			1	1
13	g12667			1	1
14	g12690			1	1
15	g12960		1		1
16	g12970			1	1
17	g13209	1	1	1	3
18	g13271			1	1
19	g13502	1	1		2
20	g13626	1			1
21	g13751			1	1
22	g13950			1	1
23	g14198			1	1
24	g14227		1	1	2
25	g1433	1		1	2
26	g14554		1	1	2
27	g14557			1	1
28	g14865			1	1
29	g14879	1			1
30	g15042	1		1	2
31	g15073	1		1	2
32	g15178			1	1
33	g15228		1	1	2
34	g15332	1			1

35	g15399			1	1
36	g15432			1	1
37	g15583		1		1
38	g15962			1	1
39	g16348			1	1
40	g16354			1	1
41	g1724		1		1
42	g1807	1			1
43	g209	1			1
44	g2242	1			1
45	g2252			1	1
46	g2333			1	1
47	g2414	1			1
48	g2531		1		1
49	g2656		1		1
50	g2669		1		1
51	g2728	1			1
52	g2978			1	1
53	g3003			1	1
54	g3035			1	1
55	g3165		1		1
56	g3168			1	1
57	g3349	1	1	1	3
58	g3354	1			1
59	g3426			1	1
60	g36	1			1
61	g3685	1			1
62	g3795			1	1
63	g3861	1			1
64	g4018	1			1
65	g4292	1			1
66	g4347		1		1
67	g4452		1		1
68	g4569	1	1		2
69	g4767			1	1
70	g4772			1	1
71	g4819		1	1	2

72	g5163			1	1
73	g5337	1	1	1	3
74	g5593	1	1	1	3
75	g5621		1		1
76	g5749			1	1
77	g5752			1	1
78	g5887	1	1	1	3
79	g6202			1	1
80	g631			1	1
81	g6350	1			1
82	g6400			1	1
83	g6612			1	1
84	g6615	1			1
85	g7033	1			1
86	g7065	1			1
87	g7262	1	1		2
88	g7281	1			1
89	g737			1	1
90	g7438			1	1
91	g7959			1	1
92	g817		1		1
93	g8330	1		1	2
94	g8348	1			1
95	g8363			1	1
96	g8683	1	1	1	3
97	g9018		1	1	2
98	g9331	1			1
99	g955	1	1	1	3
100	g9591	1		1	2
101	g978	1			1
102	g9931	1			1
103	g9943		1		1
Grand Total		42	31	63	136

Table S36: Methylation and expression levels in exons

S.NO	Scaffold	Region	Gene id	Gene description	Cytosines	Rhizine			Thallus Body			p-value	Context	Rhizine (log2RPM)	Thallus Body (log2RPM)
						Context coverage	Methylated Coverage	Methylation level	Context coverage	Methylated Coverage	Methylation level				
1	scaff30687	83050..83427	g36	urease	17	42	5	0.12	131	0	0.00	0.00	CG	4.97	5.00
2	scaff32054	27405..27812	g209	hypothetical protein EPUS_04454	20	6	2	0.33	122	2	0.02	0.01	CG	6.02	5.73
3	scaff32410	89111..89817	g955	glycoside hydrolase family 76	30	26	5	0.19	183	1	0.01	0.00	CG	7.31	7.03
4	scaff33439	18089..20638	g978	pentafunctional polypeptide	126	10	3	0.30	894	19	0.02	0.00	CG	6.33	5.99
5	scaff33055	72702..74227	g1178	transcription initiation factor TFIID subunit 1	82	35	6	0.17	573	26	0.05	0.01	CG	5.32	5.44
6	scaff31683	62703..63509	g1807	AFG3 family	45	20	4	0.20	253	14	0.06	0.03	CG	6.43	6.41
7	scaff33998	complement(57487..58433)	g2242	transcription factor tfliic complex subunit tfc6	35	45	4	0.09	357	7	0.02	0.03	CG	4.03	4.29
8	scaff33095	73878..74220	g2414	hypothetical protein VED1_08383	42	41	6	0.15	226	2	0.01	0.00	CG	6.93	6.89
9	scaff32606	complement(60427..62121)	g2728	ARM repeat	108	34	3	0.09	770	13	0.02	0.03	CG	3.84	3.78
10	scaff31957	11174..11443	g3349	hypothetical protein EMCG_04086	8	5	2	0.40	57	0	0.00	0.01	CG	3.59	4.61
11	scaff31957	24742..25327	g3354	carbohydrate-binding module family 48	40	12	3	0.25	327	3	0.01	0.00	CG	5.10	5.27
12	scaff31250	8750..9322	g3685	PIG3 family NAD(P)H quinone oxidoreductase	25	24	3	0.13	126	0	0.00	0.00	CG	6.01	6.06
13	scaff35779	34779..35550	g3861	and NB-ARC domain-containing	44	15	2	0.13	270	3	0.01	0.02	CG	4.38	4.19
14	scaff30572	26913..27528	g4018	Ebp2-domain-containing protein	29	24	3	0.13	175	4	0.02	0.04	CG	4.76	5.24
15	scaff32356	20643..22786	g4292	uncharacterized protein PAC_19922	129	47	4	0.09	933	12	0.01	0.01	CG	4.50	3.92
16	scaff32865	26883..31109	g5337	clathrin heavy chain	257	36	5	0.14	1697	70	0.04	0.02	CG	6.08	6.15
17	scaff32878	3629..4096	g5593	DUF1671 domain-containing	28	27	6	0.22	176	1	0.01	0.00	CG	6.62	6.69
18	scaff31247	31019..31492	g6350	transcription factor (Sin3)	26	10	2	0.20	160	1	0.01	0.01	CG	6.10	6.07
19	scaff32922	22978..23630	g6615	related to calcium-independent phospholipase A2	22	10	2	0.20	144	2	0.01	0.02	CG	5.48	5.67
20	scaff33137	23720..25764	g7033	oxoglutarate dehydrogenase	138	10	2	0.20	999	35	0.04	0.05	CG	6.33	6.30
21	scaff34739	14807..15616	g7065	PH-domain-containing protein	73	43	3	0.07	406	4	0.01	0.02	CG	5.40	5.67
22	scaff32455	25224..26912	g7262	hypothetical protein K441DRAFT_127390	107	10	2	0.20	822	11	0.01	0.01	CG	6.43	6.20
23	scaff37945	888..1054	g7281	ATP-dependent RNA helicase dbp2	17	106	7	0.07	61	0	0.00	0.04	CG	6.82	6.91
24	scaff32596	11873..12233	g8348	related to N-methyltransferase	25	47	4	0.09	113	1	0.01	0.03	CG	5.38	5.31
25	scaff43614	29574..30471	g8683	glycoside hydrolase family 63	66	23	4	0.17	408	13	0.03	0.01	CG	5.55	5.39
26	scaff42673	21069..21723	g9331	ARM repeat-containing	20	5	3	0.60	189	2	0.01	0.00	CG	4.10	4.05
27	scaff36025	13012..15480	g9591	MATH and UCH domain	86	5	3	0.60	588	9	0.02	0.00	CG	5.89	6.13

31	scaff52735	7430..9023	g13502	long-chain fatty-acid-ligase	88	10	2	0.20	625	14	0.02	0.02	CG	6.23	5.53
32	scaff33306	2916..3545	g13626	uncharacterized protein PAC_10190	28	25	2	0.08	158	1	0.01	0.05	CG	4.76	5.10
33	scaff44086	2070..2251	g15042	hypothetical protein V496_02323	21	27	2	0.07	181	1	0.01	0.04	CG	1.83	0.52
34	scaff41831	559..713	g15073	hypothetical protein I203_08150	14	58	24	0.41	257	68	0.26	0.02	CG	3.15	3.53
35	scaff33017	43448..43955	g1433	hypothetical protein L228DRAFT_242537	49	40	0	0.00	341	38	0.11	0.01	CG	8.18	8.34
36	scaff45925	49230..49475	g4569	ADP,ATP carrier	28	66	0	0.00	171	14	0.08	0.01	CG	8.05	7.97
37	scaff35046	29013..29436	g5887	alpha-tubulin	47	47	2	0.04	305	45	0.15	0.03	CG	7.45	7.33
38	scaff33513	11125..11583	g8330	—NA—	58	10	0	0.00	762	458	0.60	0.00	CG	9.02	9.47
39	scaff39558	11016..11186	g10979	glyceraldehyde 3-phosphate dehydrogenase	15	18	0	0.00	110	27	0.25	0.01	CG	8.29	8.25
40	scaff43241	7585..8300	g13209	alpha beta-hydrolase	37	71	0	0.00	229	15	0.07	0.02	CG	7.43	7.43
41	scaff38439	407..643	g14879	hypothetical protein EV44_g1971	11	24	4	0.17	103	47	0.46	0.01	CG	1.46	2.07
42	scaff52783	1140..1392	g15332	hypothetical protein BC1G_12337	24	101	14	0.14	754	167	0.22	0.03	CG	3.51	3.87
43	scaff30748	complement(58826..59191)	g817	Peptidyl-prolyl cis-trans isomerase-like 4	16	36	6	0.17	105	0	0.00	0.00	CHG	4.93	4.80
44	scaff32410	89111..89817	g955	glycoside hydrolase family 76	27	10	2	0.20	173	0	0.00	0.00	CHG	7.31	7.03
45	scaff30505	109925..115306	g1061	transcriptional accessory	281	7	3	0.43	1896	11	0.01	0.00	CHG	4.95	5.02
46	scaff30495	93407..98667	g1724	chitin synthase V	273	20	3	0.15	2120	57	0.03	0.02	CHG	6.52	6.56
47	scaff35492	13412..13506	g2531	cystathionine gamma-synthase	8	15	2	0.13	56	0	0.00	0.04	CHG	5.49	5.58
48	scaff32326	32450..32571	g2656	cutinase transcription factor 1 alpha	6	10	2	0.20	42	0	0.00	0.03	CHG	5.74	5.77
49	scaff32326	63336..63726	g2669	Yip1-domain-containing protein	26	21	2	0.10	233	2	0.01	0.04	CHG	6.63	6.29
50	scaff31522	32952..34445	g3165	Transcription factor steA	95	20	2	0.10	753	8	0.01	0.03	CHG	4.60	4.83
51	scaff31957	10101..11124	g3349	hypothetical protein EMCG_04086	67	10	3	0.30	482	3	0.01	0.00	CHG	3.59	4.61
52	scaff30664	34271..34700	g4347	Bifunctional solanapyrone synthase	25	5	1	0.20	160	0	0.00	0.03	CHG	5.47	5.30
53	scaff30900	8820..8837	g4452	—NA—	3	5	5	1.00	68	36	0.53	0.05	CHG	-0.96	-0.98
54	scaff31043	43148..43368	g4819	hypothetical protein IQ07DRAFT_496748	18	75	7	0.09	119	1	0.01	0.01	CHG	5.77	6.42
55	scaff32865	26883..31109	g5337	clathrin heavy chain	221	35	5	0.14	1452	61	0.04	0.02	CHG	6.08	6.15
56	scaff32878	3629..4096	g5593	DUF1671 domain-containing	28	30	6	0.20	187	2	0.01	0.00	CHG	6.62	6.69
57	scaff31300	complement(26000..26133)	g5621	anaphase-promoting complex subunit	8	12	5	0.42	40	0	0.00	0.00	CHG	3.01	3.05
58	scaff32455	25224..26912	g7262	hypothetical protein K441DRAFT_127390	84	10	2	0.20	638	17	0.03	0.03	CHG	6.43	6.20
59	scaff43614	29574..30471	g8683	glycoside hydrolase family 63	44	12	2	0.17	290	2	0.01	0.01	CHG	5.55	5.39
60	scaff32840	2030..2362	g9018	RSC complex subunit Rsc9	20	34	5	0.15	108	2	0.02	0.01	CHG	4.22	3.77
61	scaff35905	6909..7512	g9943	proline-rich protein subfamily 1	33	5	1	0.20	257	1	0.00	0.04	CHG	5.82	5.84
62	scaff37199	3288..4510	g10601	hypothetical protein LEMA_P073420.1	54	20	4	0.20	379	5	0.01	0.00	CHG	6.37	6.42
63	scaff36616	5361..8729	g11068	C2H2 and C2HC zinc finger	147	17	3	0.18	983	19	0.02	0.01	CHG	5.66	6.35
64	scaff38456	7426..7812	g12344	metallopeptidase family M24	24	10	3	0.30	153	4	0.03	0.01	CHG	5.66	5.69

65	scaff36827	1497..1597	g12960	uncharacterized protein Y057_4031	11	17	3	0.18	57	0	0.00	0.01	CHG	3.13	2.53
66	scaff52735	7430..9023	g13502	long-chain fatty-acid-ligase	83	15	3	0.20	608	20	0.03	0.01	CHG	6.23	5.53
67	scaff45248	684..711	g14227	msp1, mitochondrial	1	5	5	1.00	19	6	0.32	0.01	CHG	4.30	4.41
68	scaff34129	4503..4880	g14554	malate nad-dependent	27	25	5	0.20	178	5	0.03	0.00	CHG	7.27	7.43
69	scaff45925	49230..49475	g4569	ADP,ATP carrier	15	44	0	0.00	94	9	0.10	0.03	CHG	8.05	7.97
70	scaff35046	29013..29436	g5887	alpha-tubulin	24	26	0	0.00	148	26	0.18	0.01	CHG	7.45	7.33
71	scaff43241	7585..8300	g13209	alpha beta-hydrolase	48	105	1	0.01	291	25	0.09	0.00	CHG	7.43	7.43
72	scaff37537	1598..1693	g15228	—NA—	12	170	53	0.31	1233	512	0.42	0.01	CHG	2.35	2.43
73	scaff43629	411..468	g15583	hypothetical protein EV44_g4110	5	35	12	0.34	304	157	0.52	0.04	CHG	2.40	1.68
74	scaff30924	65669..66064	g631	hypothetical protein AYL99_06755	50	94	5	0.05	231	3	0.01	0.05	CHH	7.09	7.17
75	scaff32410	89111..89817	g955	glycoside hydrolase family 76	75	43	8	0.19	469	13	0.03	0.00	CHH	7.31	7.03
76	scaff33055	72702..74227	g1178	transcription initiation factor TFIIID subunit 1	206	35	7	0.20	1448	50	0.03	0.00	CHH	5.32	5.44
77	scaff33998	80171..80611	g2252	conserved hypothetical protein	107	40	16	0.40	790	170	0.22	0.01	CHH	8.70	8.51
78	scaff30971	61297..61678	g2978	probable autophagy (Atg22)	50	64	14	0.22	319	13	0.04	0.00	CHH	7.25	6.73
79	scaff30869	42152..42498	g3003	autophagy (Atg22)	49	20	4	0.20	328	3	0.01	0.00	CHH	6.32	6.30
80	scaff31642	37816..38400	g3035	INO80 chromatin remodeling complex	69	15	2	0.13	527	7	0.01	0.02	CHH	5.82	5.82
81	scaff31522	43934..44263	g3168	amino acid transporter	69	121	4	0.03	296	1	0.00	0.03	CHH	5.27	6.72
82	scaff31957	10101..11124	g3349	hypothetical protein EMC6_04086	182	98	4	0.04	1260	9	0.01	0.01	CHH	3.59	4.61
83	scaff32078	68390..68655	g3426	hypothetical protein K441DRAFT_680278	19	46	5	0.11	99	1	0.01	0.01	CHH	2.47	2.62
84	scaff30776	28407..29704	g4767	KH domain containing	142	40	2	0.05	958	5	0.01	0.03	CHH	5.41	5.63
85	scaff30776	39528..41175	g4772	HECT domain-containing	269	10	4	0.40	1980	30	0.02	0.00	CHH	6.49	6.64
86	scaff31043	43148..43368	g4819	hypothetical protein IQ07DRAFT_496748	52	144	14	0.10	310	2	0.01	0.00	CHH	5.77	6.42
87	scaff32396	37532..40037	g5163	ribonucleoside-diphosphate reductase large chain	319	45	3	0.07	2141	27	0.01	0.02	CHH	4.08	4.10
88	scaff32865	26883..31109	g5337	clathrin heavy chain	628	76	10	0.13	4062	168	0.04	0.00	CHH	6.08	6.15
89	scaff32878	3629..4096	g5593	DUF1671 domain-containing	99	86	16	0.19	629	9	0.01	0.00	CHH	6.62	6.69
90	scaff36119	13618..16723	g5749	initiation factor 2	390	108	7	0.06	2375	32	0.01	0.00	CHH	4.74	4.95
91	scaff36119	21065..22507	g5752	Anucleate primary sterigmata	166	137	6	0.04	1062	14	0.01	0.02	CHH	4.34	4.53
92	scaff35351	41171..42060	g6202	TorA protein	116	10	2	0.20	784	6	0.01	0.00	CHH	5.54	5.64
93	scaff31748	30851..33820	g7959	adenosinetriphosphatase	355	153	5	0.03	2550	28	0.01	0.04	CHH	5.29	5.71
94	scaff43614	29574..30471	g8683	glycoside hydrolase family 63	139	18	3	0.17	883	13	0.01	0.00	CHH	5.55	5.39
95	scaff32840	2030..2362	g9018	RSC complex subunit Rsc9	35	93	7	0.08	147	2	0.01	0.02	CHH	4.22	3.77
96	scaff36025	13012..15480	g9591	MATH and UCH domain	235	20	6	0.30	1605	11	0.01	0.00	CHH	5.89	6.13
97	scaff32942	9223..12065	g10258	viral A-type inclusion repeat	416	65	8	0.12	2730	55	0.02	0.00	CHH	5.44	5.78
98	scaff37199	3288..4510	g10601	hypothetical protein LEMA_P073420.1	153	10	2	0.20	1028	13	0.01	0.01	CHH	6.37	6.42

99	scaff38456	7426..7812	g12344	metallopeptidase family M24	50	30	6	0.20	283	11	0.04	0.00	CHH	5.66	5.69
100	scaff31268	11103..13163	g12493	dynamin-1	245	20	4	0.20	1487	16	0.01	0.00	CHH	5.81	5.79
101	scaff44988	10676..10744	g12618	hypothetical protein L228DRAFT_143430	12	73	4	0.05	93	0	0.00	0.04	CHH	4.31	4.68
102	scaff40354	4629..5921	g12634	5-3 exoribonuclease 1	188	75	4	0.05	1275	18	0.01	0.03	CHH	5.41	5.59
103	scaff33869	11690..12075	g12690	elongation factor 1-beta	72	69	24	0.35	360	87	0.24	0.05	CHH	7.72	7.70
104	scaff35811	11522..11721	g12970	glutathione S-transferase	40	20	5	0.25	364	34	0.09	0.04	CHH	7.91	7.74
105	scaff31416	4543..6616	g13751	related to splicing factor 3B subunit 3 (spliceosomal sap130)	374	59	12	0.20	2507	28	0.01	0.00	CHH	5.35	5.48
106	scaff40941	1819..3984	g14198	glycoside hydrolase family 38	339	35	4	0.11	2187	65	0.03	0.02	CHH	5.73	5.77
107	scaff45248	684..711	g14227	msp1, mitochondrial	8	31	19	0.61	168	66	0.39	0.02	CHH	4.30	4.41
108	scaff34129	4503..4880	g14554	malate nad-dependent	76	36	7	0.19	494	23	0.05	0.00	CHH	7.27	7.43
109	scaff46022	2102..2595	g14865	vacuolar sorting-associated 74	75	5	4	0.80	477	20	0.04	0.00	CHH	6.20	6.14
110	scaff44086	2070..2251	g15042	hypothetical protein V496_02323	45	67	4	0.06	378	6	0.02	0.05	CHH	1.83	0.52
111	scaff41831	559..713	g15073	hypothetical protein I203_08150	43	165	68	0.41	876	264	0.30	0.00	CHH	3.15	3.53
112	scaff51046	complement(1500..1891)	g15178	hypothetical protein PV05_09409	78	15	2	0.13	580	8	0.01	0.02	CHH	6.24	5.18
113	scaff32213	691..716	g15399	cupin domain	9	22	18	0.82	276	134	0.49	0.00	CHH	-2.99	-3.01
114	scaff36976	142..334	g15432	hypothetical protein EPUS_06399	35	15	15	1.00	259	199	0.77	0.02	CHH	-3.98	-4.00
115	scaff36268	404..634	g15962	hypothetical protein EV44_g1971	62	88	23	0.26	1599	221	0.14	0.00	CHH	2.42	3.37
116	scaff43964	672..808	g16348	—NA—	35	20	2	0.10	349	3	0.01	0.03	CHH	1.20	1.61
117	scaff30641	28459..28612	g737	—NA—	18	45	0	0.00	183	14	0.08	0.04	CHH	7.70	7.49
118	scaff33017	43448..43955	g1433	hypothetical protein L228DRAFT_242537	75	70	1	0.01	457	67	0.15	0.00	CHH	8.18	8.34
119	scaff30953	32599..33344	g2333	Auxiliary activities family 12	163	279	1	0.00	1004	26	0.03	0.01	CHH	7.75	7.58
120	scaff35731	complement(177..1819)	g3795	Glucocorticoid receptor-like (DNA-binding)	141	260	2	0.01	781	23	0.03	0.03	CHH	3.83	3.85
121	scaff35046	29013..29436	g5887	alpha-tubulin	75	53	1	0.02	467	80	0.17	0.00	CHH	7.45	7.33
122	scaff31843	4702..5016	g6400	heat shock Hsp88	71	76	2	0.03	436	45	0.10	0.02	CHH	6.88	7.04
123	scaff32922	7447..7713	g6612	Di-copper centre-containing	32	59	0	0.00	141	13	0.09	0.01	CHH	7.43	7.92
124	scaff35963	10256..10299	g7438	CMGC MAPK P38 kinase	12	63	0	0.00	29	3	0.10	0.03	CHH	7.16	7.03
125	scaff33513	11125..11583	g8330	—NA—	107	20	0	0.00	1272	799	0.63	0.00	CHH	9.02	9.47
126	scaff32960	5388..7334	g8363	cora-domain-containing protein	275	101	0	0.00	1785	56	0.03	0.04	CHH	6.19	5.86
127	scaff38190	9609..9840	g10923	gamma	46	45	0	0.00	366	35	0.10	0.01	CHH	7.94	7.81
128	scaff39558	11016..11186	g10979	glyceraldehyde 3-phosphate dehydrogenase	49	50	0	0.00	366	104	0.28	0.00	CHH	8.29	8.25
129	scaff37695	1769..1908	g11987	PNK3P-domain-containing protein	24	92	0	0.00	115	16	0.14	0.00	CHH	6.25	6.00
130	scaff37610	12411..12464	g12667	C2 domain-containing	17	102	14	0.14	441	111	0.25	0.01	CHH	1.83	1.84
131	scaff43241	7585..8300	g13209	alpha beta-hydrolase	110	160	1	0.01	707	37	0.05	0.00	CHH	7.43	7.43
132	scaff44231	3296..3398	g13271	hypothetical protein PENARI_c024G07906	31	74	28	0.38	327	166	0.51	0.03	CHH	-2.38	-2.40
133	scaff32886	2109..2193	g13950	DASH subunit Dad4	23	62	0	0.00	140	9	0.06	0.03	CHH	7.26	7.21
134	scaff45116	3558..3589	g14557	—NA—	8	25	0	0.00	52	13	0.25	0.00	CHH	-0.82	-0.49
135	scaff37537	1598..1693	g15228	—NA—	26	431	140	0.32	3016	1133	0.38	0.02	CHH	2.35	2.43
136	scaff44288	complement(381..908)	g16354	hypothetical protein V502_09950	152	549	89	0.16	3362	873	0.26	0.00	CHH	-2.61	-2.13

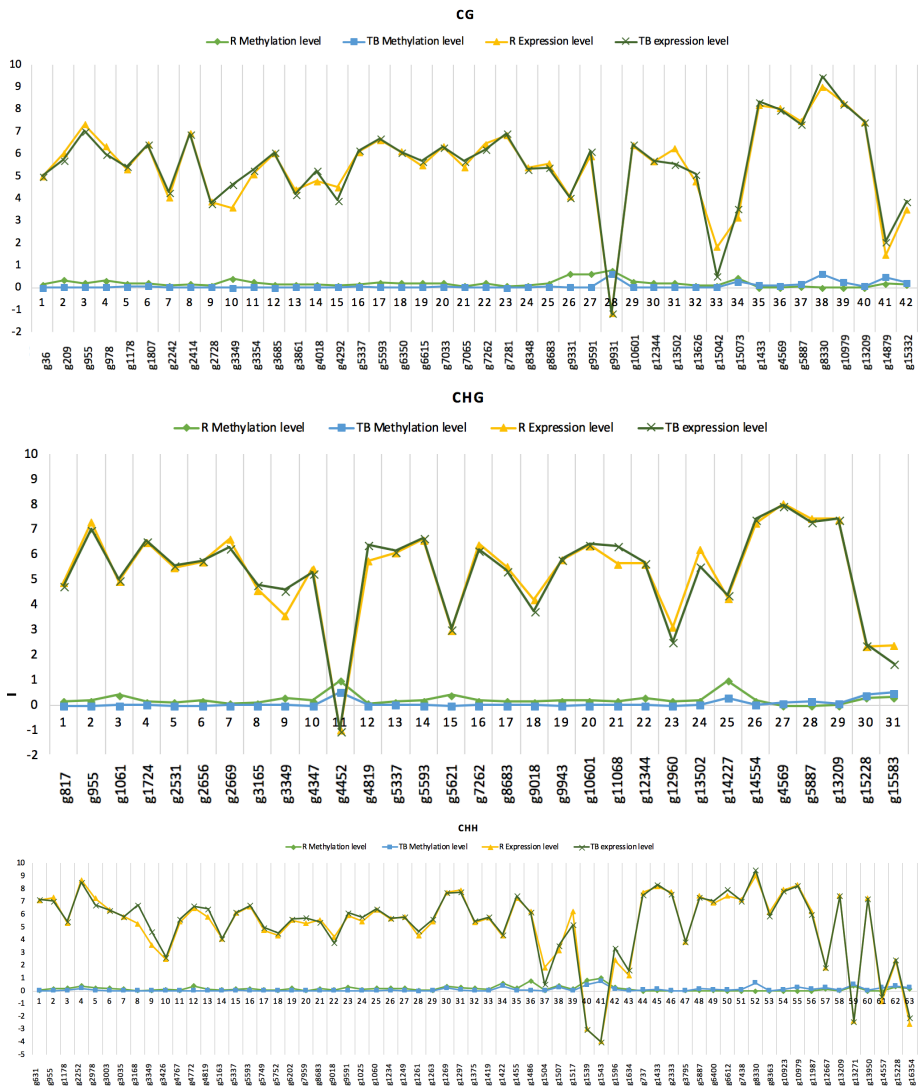


Figure S23: Exon methylation level for each sequence contexts.

Table S37: List of scaffolds and the methylated sequence contexts of differentially methylated regions for introns.

Scaffolds	CG	CHG	CHH	Grand Total
scaff30487		1	1	2
scaff30505			1	1
scaff30511	2	2	2	6
scaff30556		1	1	2
scaff30578			1	1
scaff30581			1	1
scaff30614		1		1
scaff30635			1	1
scaff30661			1	1
scaff30671		1		1
scaff30681	1	1		2
scaff30710	1			1
scaff30747		1	1	2
scaff30780			1	1
scaff30795		1	1	2
scaff30834	1			1
scaff30845		1	1	2
scaff30869			1	1
scaff30900	1	1		2
scaff30918	1			1
scaff30930	1	1	1	3
scaff30955	1	1	2	4
scaff30981		1		1
scaff31025	1	1	1	3
scaff31075			1	1
scaff31097	1	2	1	4
scaff31130	1	1		2
scaff31197			1	1
scaff31227	3	2	2	7
scaff31237		1	1	2
scaff31377	3		5	8
scaff31415		1	1	2

scaff31469			1	1
scaff31476		1	2	3
scaff31516			1	1
scaff31537	1			1
scaff31546			1	1
scaff31570			1	1
scaff31611			1	1
scaff31622	2	1	1	4
scaff31650	1			1
scaff31714	1			1
scaff31733		1		1
scaff31751	1	1	2	4
scaff31810			1	1
scaff31824	1	1	3	5
scaff31840	1		2	3
scaff31940		1		1
scaff31951			1	1
scaff31977			1	1
scaff31978		1	1	2
scaff31979			1	1
scaff31982	1	2	2	5
scaff31987			1	1
scaff32002		1		1
scaff32014	1		1	2
scaff32054			1	1
scaff32070	1	1	1	3
scaff32078			1	1
scaff32090			2	2
scaff32134			1	1
scaff32135		1		1
scaff32202			1	1
scaff32213		1	1	2
scaff32314	1	1	1	3
scaff32384		1	1	2
scaff32397			1	1
scaff32410		1		1

scaff32411		2	2
scaff32473		3	3
scaff32553	1		1
scaff32569	1		1
scaff32601	1		1
scaff32780	1	3	4
scaff32799		1	1
scaff32804	4	3	7
scaff32810	1	1	2
scaff32822	1		1
scaff32840	1		1
scaff32877	2	3	5
scaff32886		1	1
scaff32902	1	1	2
scaff32944	1		1
scaff32988		1	1
scaff33033	1		1
scaff33048	1	1	2
scaff33053	1	1	2
scaff33056		1	1
scaff33159		1	1
scaff33160		1	1
scaff33162	1	1	2
scaff33163		1	1
scaff33272		1	1
scaff33278	1		1
scaff33291		1	1
scaff33341		1	1
scaff33410		1	1
scaff33434	1	2	4
scaff33439		1	1
scaff33487	1	3	4
scaff33522		1	1
scaff33548		1	1
scaff33557	1		1
scaff33589		1	1

scaff33630		1	1	2
scaff33644		1	2	3
scaff33771		1	1	2
scaff33774	1	1		2
scaff33842			1	1
scaff33850	1		1	2
scaff33865			2	2
scaff33869			1	1
scaff33976	1			1
scaff33989	1		1	2
scaff33998	1	1	2	4
scaff34119			1	1
scaff34243	1		1	2
scaff34244			1	1
scaff34277		1	1	2
scaff34321		1	1	2
scaff34371	2		1	3
scaff34490			1	1
scaff34503			2	2
scaff34669	1	2	2	5
scaff34713			1	1
scaff34748	1	1	1	3
scaff34890			2	2
scaff35039		1	1	2
scaff35064		1	1	2
scaff35068		1	2	3
scaff35073			1	1
scaff35097		1		1
scaff35155		1	1	2
scaff35159		1		1
scaff35161		1	1	2
scaff35201			1	1
scaff35218			1	1
scaff35226	1	1		2
scaff35241		1		1
scaff35290		1	1	2

scaff35613	1		1	2
scaff35638	1		1	2
scaff35677			1	1
scaff35746			1	1
scaff35849			1	1
scaff35919		1	2	3
scaff35973	1		1	2
scaff35979			2	2
scaff36124		1	1	2
scaff36153			1	1
scaff36173		1	1	2
scaff36274			1	1
scaff36385			1	1
scaff36418		1		1
scaff36516	1			1
scaff36547			1	1
scaff36671			1	1
scaff36736	1	1	1	3
scaff36799			1	1
scaff36855			1	1
scaff36976			1	1
scaff37071	1	1	1	3
scaff37097		1	1	2
scaff37112			1	1
scaff37164	1	1	1	3
scaff37199			1	1
scaff37528	1	1	1	3
scaff37545	1			1
scaff37610	2	2	2	6
scaff37695			1	1
scaff37750	1	1	2	4
scaff37887			1	1
scaff37939	1		2	3
scaff38081			2	2
scaff38101		1	1	2
scaff38205	1		1	2

scaff38269			1	1
scaff38286			1	1
scaff38411	1	1	1	3
scaff38450	1		1	2
scaff38528	1		1	2
scaff38554	1			1
scaff38585	1		1	2
scaff38621	2	3	3	8
scaff38631	2	2	2	6
scaff38731			1	1
scaff38770		1	2	3
scaff38867			2	2
scaff38936	1			1
scaff39024	1	1	1	3
scaff39116			1	1
scaff39127	1			1
scaff39231	1			1
scaff39257			2	2
scaff39280		1		1
scaff39285	1		1	2
scaff39293		1		1
scaff39354			1	1
scaff39671			1	1
scaff40085			1	1
scaff40088			1	1
scaff40261	1	2	2	5
scaff40288		2	1	3
scaff40407			1	1
scaff40518	1	4	3	8
scaff40880			1	1
scaff41091	1	4	6	11
scaff41270	1		1	2
scaff41332			1	1
scaff41417	1			1
scaff41463	1			1
scaff41465	1			1

scaff41596			2	2
scaff41599			1	1
scaff41617			3	3
scaff41869		1		1
scaff41916			1	1
scaff41923			1	1
scaff42114			1	1
scaff42137			1	1
scaff42345			2	2
scaff42657		1	1	2
scaff42741		1		1
scaff43229	1	1	1	3
scaff43394			1	1
scaff43423		1		1
scaff43474	1	1	1	3
scaff43553			1	1
scaff43609		2	1	3
scaff43629		1	1	2
scaff43706		1	3	4
scaff43812		1	2	3
scaff43914			1	1
scaff43931	1	1	1	3
scaff44086			1	1
scaff44231	1		1	2
scaff44271			1	1
scaff44288	1			1
scaff44791	1		1	2
scaff44947	1		1	2
scaff45022	1		1	2
scaff45207			1	1
scaff45248	2	2	4	8
scaff45467	1		1	2
scaff45514			2	2
scaff45618		1		1
scaff45693		1		1
scaff46089		1	1	2

scaff46448	1	1	1	3
scaff46819			1	1
scaff46875			1	1
scaff47024			1	1
scaff47113		1	1	2
scaff47175			1	1
scaff47196			1	1
scaff47331		1	2	3
scaff47384		1	1	2
scaff47440		1		1
scaff47515			1	1
scaff47608			1	1
scaff47813		1		1
scaff48062	1			1
scaff48247			1	1
scaff48380	1			1
scaff48398	1	2	2	5
scaff48586	1			1
scaff48808		1		1
scaff49342	2		1	3
scaff49426	2	1	2	5
scaff49619		2	1	3
scaff49645		1	2	3
scaff49725	1	1		2
scaff49728			1	1
scaff49765			1	1
scaff49902		1		1
scaff49964		1	2	3
scaff50162	1			1
scaff50495			1	1
scaff51027	1		1	2
scaff51115			2	2
scaff51220		1	1	2
scaff51437		1	2	3
scaff51495			1	1
scaff51585		2	3	5

scaff51786		1		1
scaff52028		2	3	5
scaff52060	2	2	2	6
scaff52061	1	1	1	3
scaff52089	1			1
scaff52113		1		1
scaff52460			1	1
scaff52548		1	1	2
scaff52654		1	1	2
scaff52749	1	1	1	3
scaff52755	1		1	2
scaff52783	1	1	2	4
scaff52787			1	1
scaff52960	1			1
scaff53126	1	1	1	3
scaff53147	2	1	2	5
scaff53462			1	1
Grand Total	122	156	315	593

Table S38: Methylation and expression level in introns.

S.NO	Scaffolds	Region	Context	Cytosine s	Rhizine coverage	Rhizine methylation	Rhizine methylation level	Thallus Body coverage	Thallus Body methylation	Thallus Body methylation level	p-value
1	scaff31840	complement(63043..63871)	CG	30	94	27	0.29	598	107	0.18	0.01
2	scaff32902	4066..4408	CG	22	14	4	0.29	252	13	0.05	0.01
3	scaff30511	complement(55836..56159)	CG	10	77	19	0.25	537	75	0.14	0.01
4	scaff30511	70860..71981	CG	76	45	3	0.07	873	12	0.01	0.03
5	scaff31537	50270..50327	CG	3	6	3	0.50	22	0	0.00	0.01
6	scaff30900	complement(31781..32251)	CG	28	47	3	0.06	584	8	0.01	0.04
7	scaff31982	complement(54430..55128)	CG	33	346	157	0.45	1467	553	0.38	0.01
8	scaff30681	complement(3151..3650)	CG	25	10	3	0.30	207	8	0.04	0.01
9	scaff33557	13902..13952	CG	4	24	6	0.25	19	0	0.00	0.02
10	scaff35613	40655..42287	CG	71	172	69	0.40	1252	400	0.32	0.02
11	scaff37545	19311..19363	CG	3	5	3	0.60	20	0	0.00	0.00
12	scaff35973	complement(37502..37735)	CG	8	10	3	0.30	53	3	0.06	0.05
13	scaff30710	complement(39882..40123)	CG	14	1047	17	0.02	2167	17	0.01	0.03
14	scaff49725	21601..22021	CG	11	21	6	0.29	52	0	0.00	0.00
15	scaff32840	complement(1980..2029)	CG	4	12	4	0.33	20	0	0.00	0.01
16	scaff31714	complement(27743..28413)	CG	20	42	3	0.07	81	0	0.00	0.04
17	scaff32804	complement(21438..21560)	CG	5	20	12	0.60	283	105	0.37	0.04
18	scaff32804	complement(23245..23390)	CG	12	6	2	0.33	119	3	0.03	0.02
19	scaff32804	complement(25319..25391)	CG	6	5	4	0.80	105	32	0.30	0.04
20	scaff32804	26375..26445	CG	2	25	13	0.52	109	33	0.30	0.04
21	scaff32314	15753..16005	CG	14	369	177	0.48	1261	456	0.36	0.00
22	scaff31097	8029..21946	CG	582	421	56	0.13	5420	270	0.05	0.00
23	scaff39127	complement(2594..3129)	CG	23	18	2	0.11	163	1	0.01	0.03
24	scaff31025	12122..13033	CG	23	40	8	0.20	152	3	0.02	0.00
25	scaff34243	complement(12412..12615)	CG	10	206	21	0.10	829	46	0.06	0.01
26	scaff41465	1005..1077	CG	4	12	4	0.33	69	4	0.06	0.01
27	scaff33976	complement(5656..6737)	CG	62	22	3	0.14	394	8	0.02	0.02
28	scaff33162	2514..2954	CG	24	152	41	0.27	703	136	0.19	0.03
29	scaff37528	complement(1659..2610)	CG	49	192	25	0.13	1290	39	0.03	0.00
30	scaff31377	complement(2377..2594)	CG	3	66	32	0.48	329	114	0.35	0.02
31	scaff31377	2492..2632	CG	6	135	53	0.39	780	225	0.29	0.01
32	scaff31377	complement(2617..3001)	CG	34	880	361	0.41	4762	1773	0.37	0.02
33	scaff41091	2054..2246	CG	15	10	7	0.70	106	23	0.22	0.00
34	scaff37610	complement(10475..11431)	CG	53	114	53	0.46	925	266	0.29	0.00
35	scaff37610	12190..12362	CG	24	5	5	1.00	328	52	0.16	0.00
36	scaff32877	6073..6213	CG	8	113	27	0.24	447	73	0.16	0.04
37	scaff39231	336..559	CG	23	15	2	0.13	312	4	0.01	0.03
38	scaff31824	complement(3001..3412)	CG	26	135	51	0.38	1037	268	0.26	0.00
39	scaff44231	2399..2463	CG	6	20	12	0.60	101	32	0.32	0.02
40	scaff38450	4414..4612	CG	18	105	50	0.48	634	243	0.38	0.05
41	scaff33774	complement(2591..3058)	CG	24	22	11	0.50	208	30	0.14	0.00
42	scaff38528	complement(1799..1860)	CG	2	14	6	0.43	29	3	0.10	0.02
44	scaff50162	complement(4019..4404)	CG	23	5	5	1.00	308	163	0.53	0.04
45	scaff36736	complement(7575..7788)	CG	16	111	23	0.21	509	38	0.07	0.00
46	scaff31622	complement(6613..6910)	CG	24	5	2	0.40	344	10	0.03	0.01
47	scaff31622	8436..8521	CG	8	55	7	0.13	222	7	0.03	0.01
48	scaff38631	complement(3146..3325)	CG	11	103	63	0.61	482	215	0.45	0.00
49	scaff38631	complement(4497..4778)	CG	17	164	67	0.41	556	171	0.31	0.01
50	scaff45248	complement(6215..6808)	CG	55	454	202	0.44	2788	985	0.35	0.00
51	scaff31751	complement(5078..5191)	CG	10	95	7	0.07	264	6	0.02	0.03
52	scaff30834	1241..1373	CG	7	108	18	0.17	416	37	0.09	0.02
53	scaff43931	1198..1290	CG	8	26	8	0.31	79	3	0.04	0.00
54	scaff38585	complement(3349..3428)	CG	8	15	4	0.27	63	4	0.06	0.04
55	scaff39024	complement(3676..4192)	CG	17	109	32	0.29	327	35	0.11	0.00
56	scaff48062	complement(2136..2221)	CG	9	89	34	0.38	291	79	0.27	0.03
57	scaff52960	2279..2812	CG	14	19	2	0.11	169	0	0.00	0.01
58	scaff53147	469..796	CG	28	3827	1871	0.49	14595	6663	0.46	0.00
59	scaff53147	1351..1407	CG	6	460	216	0.47	1888	793	0.42	0.03
60	scaff32070	complement(378..673)	CG	25	70	32	0.46	648	172	0.27	0.00
61	scaff40261	complement(348..589)	CG	24	168	87	0.52	1377	486	0.35	0.00
62	scaff51027	complement(256..327)	CG	8	553	48	0.09	4007	261	0.07	0.04
63	scaff37071	937..981	CG	6	13	7	0.54	58	9	0.16	0.01
64	scaff48586	493..776	CG	26	74	14	0.19	442	24	0.05	0.00
65	scaff35226	complement(1588..1698)	CG	16	650	43	0.07	3168	150	0.05	0.03
66	scaff41270	complement(626..917)	CG	20	43	10	0.23	348	15	0.04	0.00
67	scaff53126	complement(1024..1182)	CG	26	173	42	0.24	1311	200	0.15	0.00
68	scaff41463	1187..1793	CG	60	150	33	0.22	1371	186	0.14	0.01
69	scaff38621	complement(255..321)	CG	8	309	114	0.37	917	222	0.24	0.00

70	scaff38621 complement(927..990)	CG	2	85	55	0.65	308	153	0.50	0.01
71	scaff49342 86..185	CG	3	78	50	0.64	376	199	0.53	0.05
72	scaff49342 328..701	CG	32	811	487	0.60	3194	1767	0.55	0.01
73	scaff52060 complement(381..632)	CG	15	7	4	0.57	184	31	0.17	0.02
74	scaff52060 405..681	CG	4	8	5	0.63	79	18	0.23	0.03
75	scaff46448 219..624	CG	45	182	58	0.32	1761	446	0.25	0.04
76	scaff38411 complement(1321..1449)	CG	11	128	5	0.04	640	7	0.01	0.04
77	scaff34371 complement(826..923)	CG	3	159	94	0.59	825	421	0.51	0.04
78	scaff34371 complement(944..1010)	CG	1	48	40	0.83	270	184	0.68	0.02
79	scaff44947 complement(601..658)	CG	4	5	3	0.60	68	4	0.06	0.01
80	scaff49426 complement(556..929)	CG	11	83	34	0.41	489	106	0.22	0.00
81	scaff45467 complement(997..1060)	CG	10	28	11	0.39	167	33	0.20	0.02
82	scaff33048 complement(155..323)	CG	9	15	9	0.60	108	25	0.23	0.01
83	scaff48380 complement(1247..1370)	CG	14	50	3	0.06	249	2	0.01	0.03
84	scaff52089 67..355	CG	19	51	38	0.75	375	190	0.51	0.00
85	scaff31227 complement(1300..1432)	CG	21	110	38	0.35	748	139	0.19	0.00
86	scaff31227 complement(1300..1453)	CG	4	34	19	0.56	164	51	0.31	0.01
87	scaff30955 457..612	CG	12	62	8	0.13	390	92	0.24	0.04
88	scaff33998 complement(85281..85630)	CG	5	20	0	0.00	5	2	0.40	0.03
89	scaff33850 complement(83239..83747)	CG	27	30	8	0.27	365	188	0.52	0.01
90	scaff41417 80340..80385	CG	4	10	0	0.00	55	18	0.33	0.03
91	scaff37750 complement(19945..20114)	CG	14	34	2	0.06	321	91	0.28	0.00
92	scaff34669 complement(4760..4894)	CG	13	62	2	0.03	87	13	0.15	0.02
93	scaff38554 complement(5592..6083)	CG	23	10	0	0.00	165	47	0.28	0.04
94	scaff31130 complement(14888..14998)	CG	15	58	1	0.02	154	15	0.10	0.04
95	scaff40518 3189..15987	CG	260	343	149	0.43	2930	1483	0.51	0.01
96	scaff33434 1082..1137	CG	8	24	0	0.00	130	27	0.21	0.01
97	scaff33989 complement(429..787)	CG	26	23	2	0.09	249	69	0.28	0.03
98	scaff32014 617..846	CG	10	16	0	0.00	138	25	0.18	0.05
99	scaff31650 9220..9838	CG	27	134	0	0.00	421	12	0.03	0.03
100	scaff37939 complement(471..536)	CG	6	52	9	0.17	395	131	0.33	0.01
101	scaff32877 6493..6556	CG	2	23	0	0.00	51	9	0.18	0.03
102	scaff30918 complement(5722..5848)	CG	14	413	26	0.06	1713	163	0.10	0.02
103	scaff39285 241..546	CG	18	5	0	0.00	239	125	0.52	0.03
104	scaff43474 complement(5950..6160)	CG	8	54	17	0.31	277	161	0.58	0.00
105	scaff44791 5940..6228	CG	36	113	2	0.02	772	49	0.06	0.03
106	scaff45248 1366..1418	CG	2	5	1	0.20	51	39	0.76	0.02
107	scaff43229 1768..1906	CG	16	285	27	0.09	1326	215	0.16	0.00
108	scaff37164 3153..3931	CG	26	6	0	0.00	208	105	0.50	0.02
109	scaff30930 complement(4761..4933)	CG	26	233	7	0.03	1168	74	0.06	0.03
110	scaff35638 complement(2787..2854)	CG	8	33	0	0.00	297	37	0.12	0.02
111	scaff34748 complement(2049..3836)	CG	110	468	17	0.04	3211	380	0.12	0.00
112	scaff52755 complement(1891..1995)	CG	10	44	2	0.05	233	38	0.16	0.03
113	scaff52783 1087..1139	CG	16	214	22	0.10	674	105	0.16	0.03
114	scaff52749 complement(106..191)	CG	6	22	3	0.14	42	27	0.64	0.00
115	scaff45022 complement(712..1432)	CG	89	990	12	0.01	707	23	0.03	0.00
116	scaff36516 701..781	CG	12	274	26	0.09	1481	230	0.16	0.00
117	scaff38205 1458..1518	CG	6	85	24	0.28	401	159	0.40	0.03
118	scaff52061 complement(1578..1852)	CG	28	281	32	0.11	1820	302	0.17	0.01
119	scaff48398 698..786	CG	6	61	21	0.34	345	188	0.54	0.00
120	scaff49426 complement(556..845)	CG	29	260	50	0.19	994	302	0.30	0.00
121	scaff44288 complement(909..960)	CG	8	27	0	0.00	165	20	0.12	0.04
122	scaff31227 complement(793..943)	CG	31	32	3	0.09	775	192	0.25	0.03
123	scaff33278 complement(120062..120956)	CHG	17	10	2	0.20	117	1	0.01	0.02
124	scaff32410 complement(87246..87761)	CHG	14	40	6	0.15	62	1	0.02	0.01
125	scaff33998 16993..17048	CHG	9	15	8	0.53	64	0	0.00	0.00
126	scaff31733 complement(1974..2027)	CHG	9	152	62	0.41	713	196	0.27	0.00
127	scaff30511 complement(55836..56159)	CHG	10	54	14	0.26	357	52	0.15	0.03
128	scaff30511 70860..71981	CHG	65	45	4	0.09	713	10	0.01	0.01
129	scaff35159 complement(65376..65462)	CHG	3	15	3	0.20	32	0	0.00	0.03
130	scaff31415 complement(1044..1544)	CHG	12	496	290	0.58	1620	836	0.52	0.00
131	scaff32384 1085..2490	CHG	58	20	9	0.45	426	47	0.11	0.00
132	scaff30900 complement(31781..32251)	CHG	30	36	5	0.14	615	18	0.03	0.01
133	scaff31982 complement(7680..8335)	CHG	4	5	3	0.60	15	0	0.00	0.01
134	scaff31982 52997..53245	CHG	20	10	7	0.70	188	29	0.15	0.00
135	scaff30671 37003..37078	CHG	10	15	3	0.20	44	0	0.00	0.01
136	scaff30681 complement(3151..3650)	CHG	22	5	2	0.40	148	2	0.01	0.01
137	scaff31476 complement(34293..34546)	CHG	24	34	13	0.38	435	77	0.18	0.01
138	scaff30614 complement(37744..37944)	CHG	17	85	6	0.07	62	0	0.00	0.03
139	scaff31940 26483..26593	CHG	3	10	3	0.30	16	0	0.00	0.05
140	scaff49725 21601..22021	CHG	13	21	6	0.29	70	1	0.01	0.00

141	scaff32822	22015..22104	CHG	14	290	6	0.02	1062	7	0.01	0.04
142	scaff35919	26271..26818	CHG	46	135	19	0.14	514	45	0.09	0.05
143	scaff31237	3527..3693	CHG	14	44	36	0.82	229	129	0.56	0.00
144	scaff36124	complement(19480..20380)	CHG	12	15	4	0.27	52	2	0.04	0.02
145	scaff33630	28213..28860	CHG	27	16	5	0.31	258	2	0.01	0.00
146	scaff33487	17357..17951	CHG	26	91	39	0.43	379	121	0.32	0.03
147	scaff33771	complement(5928..6108)	CHG	10	12	4	0.33	83	1	0.01	0.00
148	scaff33644	16140..16397	CHG	6	34	24	0.71	150	80	0.53	0.05
149	scaff39293	complement(1455..1660)	CHG	11	25	4	0.16	59	1	0.02	0.03
150	scaff32314	15753..16005	CHG	12	338	177	0.52	1085	449	0.41	0.00
151	scaff31097	8029..21946	CHG	589	420	87	0.21	5289	407	0.08	0.00
152	scaff30747	complement(7446..7491)	CHG	3	9	6	0.67	19	3	0.16	0.01
153	scaff38101	9527..9691	CHG	14	96	14	0.15	429	34	0.08	0.04
154	scaff31025	12122..13033	CHG	14	55	6	0.11	114	0	0.00	0.00
155	scaff34321	complement(1643..1694)	CHG	6	17	10	0.59	145	24	0.17	0.00
156	scaff40518	complement(18268..18489)	CHG	10	35	22	0.63	163	60	0.37	0.00
157	scaff35241	1297..1879	CHG	56	596	369	0.62	3439	1962	0.57	0.01
158	scaff37528	complement(1659..2610)	CHG	51	249	22	0.09	1453	60	0.04	0.00
159	scaff33033	complement(3979..4257)	CHG	18	35	4	0.11	145	3	0.02	0.03
160	scaff32601	2868..2923	CHG	3	10	3	0.30	28	1	0.04	0.05
161	scaff35155	11384..11458	CHG	2	22	6	0.27	36	2	0.06	0.03
162	scaff35097	complement(579..764)	CHG	14	43	3	0.07	109	0	0.00	0.02
163	scaff41091	314..384	CHG	3	16	8	0.50	40	8	0.20	0.03
164	scaff41091	1255..1312	CHG	3	11	7	0.64	37	9	0.24	0.02
165	scaff37610	complement(10475..11431)	CHG	61	230	111	0.48	1347	504	0.37	0.00
166	scaff37610	12190..12362	CHG	26	17	8	0.47	382	73	0.19	0.01
167	scaff35068	2835..3158	CHG	17	29	19	0.66	298	88	0.30	0.00
168	scaff31824	complement(3001..3412)	CHG	31	177	95	0.54	1489	645	0.43	0.01
169	scaff45693	complement(5901..6143)	CHG	25	40	4	0.10	401	4	0.01	0.00
170	scaff32002	6689..6877	CHG	4	220	145	0.66	605	354	0.59	0.03
171	scaff33774	complement(2591..3058)	CHG	32	44	21	0.48	321	104	0.32	0.03
172	scaff36736	complement(7575..7788)	CHG	10	76	16	0.21	499	60	0.12	0.03
173	scaff37097	1407..1897	CHG	13	5	4	0.80	103	27	0.26	0.02
174	scaff35290	complement(6505..7586)	CHG	64	76	22	0.29	691	93	0.13	0.00
175	scaff31622	8436..8521	CHG	5	46	7	0.15	158	4	0.03	0.00
176	scaff38631	complement(3146..3325)	CHG	11	109	76	0.70	519	288	0.55	0.00
177	scaff38631	complement(4497..4778)	CHG	20	222	118	0.53	744	346	0.47	0.05
178	scaff45248	2900..2988	CHG	7	17	9	0.53	42	4	0.10	0.00
179	scaff45248	complement(6215..6808)	CHG	50	562	292	0.52	2861	1316	0.46	0.01
180	scaff41869	4317..4832	CHG	23	10	2	0.20	145	2	0.01	0.02
181	scaff30981	5642..6147	CHG	15	28	3	0.11	290	1	0.00	0.00
182	scaff49964	complement(4105..4177)	CHG	8	88	27	0.31	286	45	0.16	0.00
183	scaff43423	complement(4127..4910)	CHG	79	732	350	0.48	3797	1663	0.44	0.03
184	scaff47384	complement(4064..4570)	CHG	21	80	46	0.58	564	254	0.45	0.02
185	scaff43931	1198..1290	CHG	4	15	10	0.67	30	5	0.17	0.00
186	scaff30930	5245..5302	CHG	3	5	4	0.80	43	0	0.00	0.00
187	scaff39024	complement(3676..4192)	CHG	32	279	111	0.40	934	148	0.16	0.00
188	scaff52028	2921..3130	CHG	17	77	21	0.27	449	82	0.18	0.05
189	scaff32944	528..768	CHG	15	69	42	0.61	400	158	0.40	0.00
190	scaff43812	892..1351	CHG	49	420	233	0.55	2055	958	0.47	0.00
191	scaff47440	932..1179	CHG	17	105	5	0.05	565	8	0.01	0.04
192	scaff43706	2731..2801	CHG	8	524	376	0.72	2988	1933	0.65	0.00
193	scaff32780	complement(2016..2900)	CHG	45	295	161	0.55	1257	596	0.47	0.02
194	scaff33053	109..335	CHG	11	46	28	0.61	217	79	0.36	0.00
195	scaff32810	2273..2739	CHG	22	64	16	0.25	421	55	0.13	0.01
196	scaff45618	1395..1447	CHG	5	9	5	0.56	80	12	0.15	0.01
197	scaff51585	complement(1321..1461)	CHG	4	17	17	1.00	83	60	0.72	0.01
198	scaff51585	1949..2003	CHG	3	111	86	0.77	332	212	0.64	0.00
199	scaff53147	469..796	CHG	25	2708	1757	0.65	10648	6543	0.61	0.00
200	scaff32070	complement(378..673)	CHG	16	34	16	0.47	501	151	0.30	0.03
201	scaff46089	complement(250..475)	CHG	10	16	6	0.38	148	23	0.16	0.04
202	scaff47113	2043..2098	CHG	5	144	48	0.33	568	147	0.26	0.05
203	scaff32213	complement(1285..1860)	CHG	23	43	34	0.79	516	161	0.31	0.00
204	scaff40261	complement(348..408)	CHG	8	70	40	0.57	299	120	0.40	0.01
205	scaff40261	complement(348..589)	CHG	18	163	107	0.66	1271	656	0.52	0.00
206	scaff39280	1536..1592	CHG	7	5	2	0.40	78	1	0.01	0.01
207	scaff37071	937..981	CHG	5	12	8	0.67	55	13	0.24	0.01
208	scaff52749	complement(874..983)	CHG	12	43	12	0.28	163	24	0.15	0.04
209	scaff35226	complement(1588..1698)	CHG	14	570	38	0.07	2735	114	0.04	0.01
210	scaff43629	complement(607..694)	CHG	10	101	54	0.53	801	345	0.43	0.03
211	scaff53126	complement(1024..1182)	CHG	12	81	21	0.26	634	109	0.17	0.04
212	scaff42741	complement(576..927)	CHG	6	20	3	0.15	59	0	0.00	0.01
213	scaff40288	911..975	CHG	13	14	12	0.86	189	72	0.38	0.00
214	scaff38621	complement(255..321)	CHG	9	283	139	0.49	855	284	0.33	0.00
215	scaff38621	505..569	CHG	8	195	132	0.68	1146	697	0.61	0.04
216	scaff38621	complement(927..990)	CHG	4	121	80	0.66	475	261	0.55	0.02

217	scaff51437	370..444	CHG	17	52	25	0.48	446	105	0.24	0.00
218	scaff52060	complement(381..632)	CHG	40	49	19	0.39	514	128	0.25	0.03
219	scaff52060	405..681	CHG	9	28	15	0.54	173	31	0.18	0.00
220	scaff46448	219..624	CHG	39	133	70	0.53	1629	599	0.37	0.00
221	scaff43609	complement(295..460)	CHG	15	616	392	0.64	2447	1440	0.59	0.02
222	scaff38411	complement(864..1252)	CHG	12	22	3	0.14	75	0	0.00	0.01
223	scaff32135	complement(640..971)	CHG	29	125	80	0.64	1106	596	0.54	0.02
224	scaff49645	complement(1101..1143)	CHG	7	49	15	0.31	205	36	0.18	0.04
225	scaff52548	435..639	CHG	22	235	133	0.57	1019	461	0.45	0.00
226	scaff49619	1460..1509	CHG	5	48	24	0.50	194	59	0.30	0.01
227	scaff49619	1460..1520	CHG	2	15	11	0.73	71	26	0.37	0.01
228	scaff31227	complement(1300..1453)	CHG	4	21	12	0.57	138	36	0.26	0.01
229	scaff30955	457..612	CHG	10	67	7	0.10	320	66	0.21	0.03
230	scaff32569	complement(79653..79768)	CHG	7	216	24	0.11	620	112	0.18	0.01
231	scaff37750	complement(19291..19873)	CHG	16	12	2	0.17	137	65	0.47	0.04
232	scaff31978	37289..37357	CHG	6	22	5	0.23	104	47	0.45	0.04
233	scaff34669	complement(4760..4894)	CHG	8	38	2	0.05	74	14	0.19	0.04
234	scaff34669	complement(5792..5831)	CHG	1	9	1	0.11	33	16	0.48	0.05
235	scaff30795	complement(3614..3948)	CHG	28	35	1	0.03	206	32	0.16	0.03
236	scaff35161	4010..4371	CHG	18	12	3	0.25	187	137	0.73	0.00
237	scaff31097	24079..24545	CHG	19	19	0	0.00	174	46	0.26	0.00
238	scaff30556	17353..17511	CHG	4	13	4	0.31	53	35	0.66	0.02
239	scaff31130	complement(14888..14998)	CHG	12	48	2	0.04	128	20	0.16	0.03
240	scaff40518	3189..15987	CHG	421	391	187	0.48	4544	2518	0.55	0.00
241	scaff40518	16790..16906	CHG	4	8	1	0.13	42	24	0.57	0.02
242	scaff40518	20622..20691	CHG	6	6	0	0.00	67	36	0.54	0.01
243	scaff35064	complement(5766..5819)	CHG	7	44	0	0.00	214	17	0.08	0.04
244	scaff33434	1018..1079	CHG	5	26	0	0.00	165	28	0.17	0.01
245	scaff33434	1082..1137	CHG	8	24	0	0.00	119	27	0.23	0.00
246	scaff32553	complement(16450..16664)	CHG	4	7	0	0.00	66	37	0.56	0.01
247	scaff34277	complement(9665..10274)	CHG	22	72	4	0.06	288	38	0.13	0.05
248	scaff41091	2912..2985	CHG	6	21	2	0.10	199	56	0.28	0.05
249	scaff41091	complement(3754..4166)	CHG	31	15	0	0.00	222	46	0.21	0.04
250	scaff35039	4616..4798	CHG	10	38	4	0.11	179	43	0.24	0.05
251	scaff47813	complement(9575..10060)	CHG	22	261	156	0.60	1012	674	0.67	0.02
252	scaff30487	complement(2314..2399)	CHG	6	98	2	0.02	366	39	0.11	0.00
253	scaff43474	complement(5950..6160)	CHG	18	160	87	0.54	646	407	0.63	0.03
254	scaff43229	1768..1906	CHG	3	95	15	0.16	377	92	0.24	0.05
255	scaff31751	4834..5063	CHG	15	47	0	0.00	215	15	0.07	0.05
256	scaff37164	3153..3931	CHG	29	24	3	0.13	226	150	0.66	0.00
257	scaff52028	complement(2777..2855)	CHG	10	16	0	0.00	194	62	0.32	0.00
258	scaff49902	599..1505	CHG	53	145	6	0.04	1032	86	0.08	0.05
259	scaff30845	complement(1258..1339)	CHG	14	20	0	0.00	243	42	0.17	0.03
260	scaff34748	complement(2049..3836)	CHG	110	409	25	0.06	2987	426	0.14	0.00
261	scaff52113	complement(396..434)	CHG	5	47	14	0.30	109	51	0.47	0.03
262	scaff36418	2100..2322	CHG	15	1721	57	0.03	6627	311	0.05	0.01
263	scaff42657	1014..1059	CHG	4	38	2	0.05	194	41	0.21	0.01
264	scaff38770	1307..1721	CHG	35	127	39	0.31	627	248	0.40	0.04
265	scaff52783	746..861	CHG	15	81	16	0.20	530	182	0.34	0.01
266	scaff51220	833..879	CHG	8	138	45	0.33	674	300	0.45	0.01
267	scaff40288	911..981	CHG	2	6	0	0.00	25	15	0.60	0.01
268	scaff47331	complement(1573..1609)	CHG	7	270	95	0.35	1160	505	0.44	0.01
269	scaff52654	complement(644..722)	CHG	13	98	15	0.15	540	156	0.29	0.00
270	scaff48808	505..572	CHG	7	199	75	0.38	1025	478	0.47	0.01
271	scaff43609	complement(668..811)	CHG	7	42	15	0.36	189	98	0.52	0.04
272	scaff36173	695..819	CHG	9	31	11	0.35	152	92	0.61	0.01
273	scaff52061	complement(1578..1852)	CHG	13	133	14	0.11	813	175	0.22	0.00
274	scaff48398	complement(228..292)	CHG	12	32	9	0.28	316	158	0.50	0.01
275	scaff48398	698..786	CHG	9	95	33	0.35	550	277	0.50	0.00
276	scaff49426	complement(556..845)	CHG	22	200	57	0.29	729	262	0.36	0.03
277	scaff51786	complement(624..675)	CHG	7	43	19	0.44	285	182	0.64	0.01
278	scaff31227	complement(793..943)	CHG	7	17	2	0.12	186	63	0.34	0.05
279	scaff32054	22928..23356	CHH	12	15	2	0.13	59	0	0.00	0.04
280	scaff31570	complement(43273..43330)	CHH	12	25	3	0.12	136	1	0.01	0.01
281	scaff30505	complement(1441..1495)	CHH	6	15	8	0.53	91	18	0.20	0.01
282	scaff31840	complement(63043..63871)	CHH	137	289	103	0.36	1640	493	0.30	0.04
283	scaff33998	16993..17048	CHH	15	25	8	0.32	110	1	0.01	0.00
284	scaff35677	complement(88424..88511)	CHH	23	38	18	0.47	767	220	0.29	0.01
285	scaff32902	4066..4408	CHH	49	21	5	0.24	536	22	0.04	0.00

286	scaff33163	49571..50225	CHH	160	112	86	0.77	1646	973	0.59	0.00
287	scaff30869	48774..48835	CHH	15	42	6	0.14	75	0	0.00	0.00
288	scaff34119	complement(8344..8510)	CHH	39	87	29	0.33	409	50	0.12	0.00
289	scaff32078	complement(68466..68524)	CHH	12	46	5	0.11	55	0	0.00	0.02
290	scaff330511	complement(55836..56159)	CHH	35	199	58	0.29	1263	201	0.16	0.00
291	scaff330511	70860..71981	CHH	212	127	20	0.16	2097	29	0.01	0.00
292	scaff33056	complement(7499..7691)	CHH	29	94	6	0.06	129	1	0.01	0.02
293	scaff31415	complement(1044..1544)	CHH	53	2225	1123	0.50	7345	3372	0.46	0.00
294	scaff32384	1085..2490	CHH	195	60	19	0.32	1480	159	0.11	0.00
295	scaff31197	complement(61510..62284)	CHH	36	45	6	0.13	174	2	0.01	0.00
296	scaff39116	28958..29119	CHH	19	434	9	0.02	948	7	0.01	0.03
297	scaff31469	29907..30014	CHH	18	25	3	0.12	82	1	0.01	0.04
298	scaff33159	complement(28964..35644)	CHH	1727	7321	3819	0.52	39547	19583	0.50	0.00
299	scaff35613	40655..42287	CHH	331	818	323	0.39	5337	1623	0.30	0.00
300	scaff31476	complement(34293..34546)	CHH	64	188	63	0.34	1632	299	0.18	0.00
301	scaff33341	12063..13146	CHH	125	10	3	0.30	1006	29	0.03	0.00
302	scaff31546	23702..23811	CHH	28	16	2	0.13	259	2	0.01	0.02
303	scaff36855	complement(18335..18444)	CHH	14	15	2	0.13	130	0	0.00	0.01
304	scaff35919	26271..26818	CHH	77	297	36	0.12	874	74	0.08	0.04
305	scaff31237	3527..3693	CHH	32	35	27	0.77	452	230	0.51	0.00
306	scaff36124	complement(19480..20380)	CHH	44	35	4	0.11	202	1	0.00	0.00
307	scaff31075	5567..5897	CHH	27	54	4	0.07	221	3	0.01	0.03
308	scaff36671	complement(11750..12995)	CHH	125	48	4	0.08	1022	17	0.02	0.01
309	scaff40085	1051..1183	CHH	35	492	117	0.24	1244	172	0.14	0.00
310	scaff39671	complement(114..195)	CHH	28	164	88	0.54	335	150	0.45	0.04
311	scaff33630	28213..28860	CHH	95	99	5	0.05	852	8	0.01	0.01
312	scaff33487	complement(16098..16251)	CHH	26	10	3	0.30	192	13	0.07	0.03
313	scaff33487	17357..17951	CHH	140	466	200	0.43	2087	771	0.37	0.01
314	scaff33487	18036..18697	CHH	129	1177	542	0.46	4590	1865	0.41	0.00
315	scaff30661	5294..6276	CHH	106	176	12	0.07	638	7	0.01	0.00
316	scaff33771	complement(5928..6108)	CHH	20	26	8	0.31	168	1	0.01	0.00
317	scaff34503	10556..10868	CHH	11	77	43	0.56	227	98	0.43	0.04
318	scaff32411	478..547	CHH	22	259	162	0.63	819	425	0.52	0.00
319	scaff32411	12513..12826	CHH	86	472	228	0.48	2188	889	0.41	0.00
320	scaff33644	16140..16353	CHH	41	7	6	0.86	345	82	0.24	0.00
321	scaff33644	16140..16397	CHH	12	81	42	0.52	314	127	0.40	0.04
322	scaff32314	15753..16005	CHH	78	1455	716	0.49	4965	2099	0.42	0.00
323	scaff31516	22588..22678	CHH	20	199	128	0.64	791	447	0.57	0.03
324	scaff31951	complement(9434..9809)	CHH	65	43	5	0.12	667	14	0.02	0.00
325	scaff31097	8029..21946	CHH	2065	1902	382	0.20	19239	1566	0.08	0.00
326	scaff35979	complement(18035..18188)	CHH	37	30	23	0.77	621	331	0.53	0.01
327	scaff34890	22238..22600	CHH	116	1566	760	0.49	6801	3119	0.46	0.03
328	scaff34890	22238..22624	CHH	9	125	93	0.74	446	277	0.62	0.01
329	scaff30747	complement(7446..7491)	CHH	9	28	22	0.79	57	10	0.18	0.00
330	scaff38101	9527..9691	CHH	41	316	54	0.17	1153	139	0.12	0.01
331	scaff31025	12122..13033	CHH	84	304	35	0.12	631	17	0.03	0.00
332	scaff34321	complement(1643..1694)	CHH	18	62	29	0.47	467	80	0.17	0.00
333	scaff34243	complement(12412..12615)	CHH	21	356	44	0.12	1892	162	0.09	0.02
334	scaff40518	21126..21228	CHH	23	10	8	0.80	211	102	0.48	0.05
335	scaff33865	16856..16902	CHH	12	40	16	0.40	242	56	0.23	0.02
336	scaff33865	17304..17397	CHH	30	62	31	0.50	528	147	0.28	0.00
337	scaff33162	2514..2954	CHH	100	1054	506	0.48	3822	1713	0.45	0.04
338	scaff37528	complement(1659..2610)	CHH	148	666	68	0.10	3929	207	0.05	0.00
339	scaff38867	complement(17006..17064)	CHH	2	7	5	0.71	40	12	0.30	0.05
340	scaff31377	1382..1456	CHH	24	86	36	0.42	627	164	0.26	0.00
341	scaff31377	complement(2377..2576)	CHH	53	510	214	0.42	4097	1468	0.36	0.00
342	scaff31377	complement(2377..2594)	CHH	6	98	55	0.56	650	296	0.46	0.03
343	scaff31377	2492..2632	CHH	8	155	90	0.58	1076	427	0.40	0.00
344	scaff35155	11384..11458	CHH	5	33	12	0.36	77	6	0.08	0.00
345	scaff36274	complement(301..345)	CHH	11	221	131	0.59	566	262	0.46	0.00
346	scaff31979	6054..6179	CHH	30	86	5	0.06	253	2	0.01	0.01
347	scaff37695	2362..2431	CHH	12	43	3	0.07	79	0	0.00	0.04
348	scaff33410	complement(7007..7740)	CHH	112	129	69	0.53	924	49	0.05	0.00
349	scaff32473	10801..10858	CHH	10	131	62	0.47	624	236	0.38	0.03
350	scaff41916	11618..11833	CHH	41	102	6	0.06	522	4	0.01	0.00
351	scaff38731	10337..10411	CHH	17	5	2	0.40	130	1	0.01	0.00
352	scaff41091	314..384	CHH	12	36	17	0.47	142	24	0.17	0.00
353	scaff41091	1215..1263	CHH	17	72	26	0.36	289	68	0.24	0.02
354	scaff41091	2054..2246	CHH	45	5	4	0.80	341	70	0.21	0.01
355	scaff37610	complement(10475..11431)	CHH	294	951	454	0.48	5236	1807	0.35	0.00
356	scaff37610	12190..12362	CHH	28	35	16	0.46	439	96	0.22	0.00
357	scaff33869	674..753	CHH	20	7	2	0.29	115	1	0.01	0.01
358	scaff32877	6073..6213	CHH	32	528	145	0.27	1928	424	0.22	0.01
359	scaff32877	10328..10930	CHH	157	775	51	0.07	3536	50	0.01	0.00
360	scaff35068	2835..3158	CHH	88	132	64	0.48	1565	458	0.29	0.00

361	scaff34490 complement(612..1499)	CHH	79	35	8	0.23	549	6	0.01	0.00
362	scaff31824 complement(3001..3412)	CHH	104	510	203	0.40	3270	1172	0.36	0.05
363	scaff31824 complement(3001..3433)	CHH	11	227	132	0.58	1248	607	0.49	0.01
364	scaff34244 11236..11797	CHH	68	52	3	0.06	408	3	0.01	0.02
365	scaff44231 3226..3278	CHH	10	97	45	0.46	208	66	0.32	0.01
366	scaff38450 4414..4612	CHH	51	356	175	0.49	2031	879	0.43	0.02
367	scaff33160 complement(7715..8278)	CHH	133	394	139	0.35	2749	700	0.25	0.00
368	scaff33589 5248..5427	CHH	15	10	10	1.00	155	74	0.48	0.00
369	scaff38528 complement(1799..1860)	CHH	18	199	71	0.36	348	73	0.21	0.00
370	scaff30635 complement(695..757)	CHH	18	27	23	0.85	153	63	0.41	0.00
371	scaff32799 5881..5987	CHH	23	20	11	0.55	349	71	0.20	0.00
372	scaff36736 complement(7575..7788)	CHH	41	206	40	0.19	1226	155	0.13	0.01
373	scaff37097 1407..1897	CHH	42	49	24	0.49	408	117	0.29	0.00
374	scaff35290 complement(6505..7586)	CHH	255	399	149	0.37	2513	472	0.19	0.00
375	scaff31622 8436..8521	CHH	16	88	15	0.17	382	16	0.04	0.00
376	scaff33548 complement(4099..4178)	CHH	26	673	413	0.61	2878	1659	0.58	0.04
377	scaff47608 5779..5962	CHH	26	93	25	0.27	229	23	0.10	0.00
378	scaff38631 complement(3146..3325)	CHH	58	733	429	0.59	3309	1596	0.48	0.00
379	scaff38631 complement(5790..5857)	CHH	21	1000	556	0.56	4166	2144	0.51	0.01
380	scaff30780 5034..5331	CHH	26	20	3	0.15	194	1	0.01	0.00
381	scaff32134 4528..4653	CHH	21	17	10	0.59	335	88	0.26	0.01
382	scaff45248 2900..2988	CHH	22	49	24	0.49	133	13	0.10	0.00
383	scaff45248 5778..5819	CHH	10	55	19	0.35	264	57	0.22	0.03
384	scaff45248 complement(6215..6808)	CHH	168	2044	1078	0.53	10046	4886	0.49	0.00
385	scaff49964 481..667	CHH	59	355	165	0.46	1397	537	0.38	0.00
386	scaff49964 complement(4105..4177)	CHH	21	258	86	0.33	668	143	0.21	0.00
387	scaff31751 complement(5078..5191)	CHH	28	191	10	0.05	609	14	0.02	0.04
388	scaff47384 complement(4064..4570)	CHH	111	508	241	0.47	3169	1248	0.39	0.00
389	scaff43931 1198..1290	CHH	21	87	30	0.34	180	20	0.11	0.00
390	scaff38585 complement(3349..3428)	CHH	12	10	4	0.40	92	10	0.11	0.03
391	scaff42137 1692..1758	CHH	23	59	6	0.10	557	14	0.03	0.01
392	scaff39024 complement(3676..4192)	CHH	108	944	263	0.28	3633	503	0.14	0.00
393	scaff31987 1638..1842	CHH	59	224	41	0.18	817	28	0.03	0.00
394	scaff33522 complement(3422..3494)	CHH	12	7	1	0.14	137	0	0.00	0.05
395	scaff52028 2921..3130	CHH	64	438	141	0.32	2441	653	0.27	0.01
396	scaff52028 3316..3380	CHH	24	60	21	0.35	304	71	0.23	0.04
397	scaff41599 1483..1584	CHH	23	34	10	0.29	423	59	0.14	0.02
398	scaff43812 892..1351	CHH	136	1384	602	0.43	6695	2607	0.39	0.00
399	scaff43812 1516..1706	CHH	48	214	90	0.42	1139	326	0.29	0.00
400	scaff43706 2731..2801	CHH	23	1595	968	0.61	9157	4962	0.54	0.00
401	scaff43706 2822..2907	CHH	34	2522	1417	0.56	14067	7376	0.52	0.00
402	scaff43706 2822..2919	CHH	2	98	74	0.76	518	274	0.53	0.00
403	scaff32780 complement(2016..2900)	CHH	175	1310	729	0.56	4780	2299	0.48	0.00
404	scaff32780 complement(2016..2915)	CHH	6	157	84	0.54	371	157	0.42	0.01
405	scaff33053 109..335	CHH	56	370	168	0.45	1339	513	0.38	0.01
406	scaff49765 complement(1699..1787)	CHH	20	262	45	0.17	1470	149	0.10	0.00
407	scaff32810 983..1329	CHH	84	61	3	0.05	1439	17	0.01	0.04
408	scaff44086 211..474	CHH	40	36	11	0.31	371	49	0.13	0.01
409	scaff42345 1255..1799	CHH	122	441	117	0.27	1936	406	0.21	0.01
410	scaff42345 1615..1820	CHH	6	10	5	0.50	97	19	0.20	0.04
411	scaff35849 complement(1226..1291)	CHH	22	60	26	0.43	665	157	0.24	0.00
412	scaff36547 complement(2450..2503)	CHH	17	118	31	0.26	717	63	0.09	0.00
413	scaff43553 1400..1462	CHH	21	19	12	0.63	439	160	0.36	0.02
414	scaff51585 805..1434	CHH	10	10	10	1.00	158	92	0.58	0.01
415	scaff51585 complement(1321..1461)	CHH	7	6	6	1.00	172	71	0.41	0.01
416	scaff51585 1949..2003	CHH	14	383	207	0.54	1369	619	0.45	0.00
417	scaff53147 469..796	CHH	116	16271	9138	0.56	62146	33574	0.54	0.00
418	scaff53147 1351..1407	CHH	16	1268	749	0.59	5430	2662	0.49	0.00
419	scaff46819 complement(802..1035)	CHH	64	117	44	0.38	866	157	0.18	0.00
420	scaff32070 complement(378..673)	CHH	79	214	88	0.41	2146	482	0.22	0.00
421	scaff46089 complement(250..475)	CHH	56	65	39	0.60	782	180	0.23	0.00
422	scaff47113 2043..2098	CHH	20	495	144	0.29	1794	436	0.24	0.02
423	scaff32213 complement(1285..1860)	CHH	71	169	77	0.46	1981	667	0.34	0.00
424	scaff49728 400..507	CHH	15	50	7	0.14	694	17	0.02	0.00
425	scaff40407 complement(1154..1292)	CHH	19	65	23	0.35	220	52	0.24	0.04
426	scaff40261 complement(348..408)	CHH	18	75	39	0.52	641	220	0.34	0.00
427	scaff40261 complement(348..589)	CHH	56	470	284	0.60	2345	1111	0.47	0.00
428	scaff51027 complement(256..327)	CHH	20	1422	127	0.09	9738	735	0.08	0.04
429	scaff30581 1402..1690	CHH	91	229	81	0.35	821	122	0.15	0.00
430	scaff38286 complement(125..203)	CHH	27	194	104	0.54	1280	570	0.45	0.01
431	scaff52749 complement(874..983)	CHH	35	90	24	0.27	448	70	0.16	0.01
432	scaff47515 complement(252..314)	CHH	20	68	34	0.50	594	225	0.38	0.04
433	scaff43629 complement(607..694)	CHH	29	242	130	0.54	2269	998	0.44	0.00
434	scaff41270 complement(626..917)	CHH	92	163	29	0.18	1622	147	0.09	0.00
435	scaff40880 complement(173..254)	CHH	16	84	12	0.14	202	12	0.06	0.02
436	scaff41617 complement(589..643)	CHH	17	39	14	0.36	312	52	0.17	0.01
437	scaff41617 complement(751..835)	CHH	20	109	38	0.35	473	83	0.18	0.00
438	scaff41596 complement(425..1062)	CHH	202	748	244	0.33	6170	1755	0.28	0.01
439	scaff41596 complement(1107..1206)	CHH	33	42	22	0.52	894	293	0.33	0.01
440	scaff53462 complement(201..642)	CHH	65	31	8	0.26	655	79	0.12	0.03

441	scaff53126 complement(1024..1182)	CHH	43	360	96	0.27	2391	379	0.16	0.00
442	scaff40288 911..975	CHH	18	14	12	0.86	261	111	0.43	0.00
443	scaff38621 complement(255..321)	CHH	16	659	257	0.39	1831	592	0.32	0.00
444	scaff38621 505..569	CHH	22	588	307	0.52	3295	1575	0.48	0.03
445	scaff38621 505..593	CHH	8	206	135	0.66	1400	790	0.56	0.01
446	scaff36385 821..1029	CHH	44	80	39	0.49	621	163	0.26	0.00
447	scaff49342 86..185	CHH	38	853	571	0.67	4627	2838	0.61	0.00
448	scaff51115 692..882	CHH	58	498	216	0.43	1969	753	0.38	0.02
449	scaff51437 370..444	CHH	23	98	40	0.41	671	185	0.28	0.01
450	scaff38269 complement(942..1144)	CHH	54	356	161	0.45	2220	748	0.34	0.00
451	scaff52060 complement(381..632)	CHH	75	114	42	0.37	974	183	0.19	0.00
452	scaff52060 405..681	CHH	14	22	10	0.45	252	38	0.15	0.00
453	scaff52460 246..350	CHH	19	40	8	0.20	247	22	0.09	0.04
454	scaff46448 219..624	CHH	109	549	261	0.48	4773	1724	0.36	0.00
455	scaff43609 complement(295..460)	CHH	68	3135	1769	0.56	13083	7098	0.54	0.01
456	scaff35073 complement(1016..1170)	CHH	60	112	54	0.48	973	343	0.35	0.01
457	scaff38411 complement(864..1252)	CHH	45	107	7	0.07	235	2	0.01	0.01
458	scaff41923 1247..1337	CHH	26	366	163	0.45	1657	645	0.39	0.03
459	scaff34371 complement(826..923)	CHH	38	1991	1130	0.57	10949	5776	0.53	0.00
460	scaff44947 complement(601..658)	CHH	16	62	30	0.48	301	50	0.17	0.00
461	scaff33272 440..504	CHH	20	88	46	0.52	420	139	0.33	0.00
462	scaff49426 complement(556..929)	CHH	21	130	47	0.36	980	235	0.24	0.00
463	scaff45467 complement(997..1060)	CHH	18	59	26	0.44	322	66	0.20	0.00
464	scaff49645 complement(1101..1143)	CHH	11	36	9	0.25	181	21	0.12	0.04
465	scaff43914 complement(377..440)	CHH	14	27	21	0.78	502	276	0.55	0.01
466	scaff33048 complement(155..323)	CHH	41	37	15	0.41	474	119	0.25	0.03
467	scaff52548 435..639	CHH	72	806	430	0.53	3185	1328	0.42	0.00
468	scaff49619 1460..1509	CHH	13	136	92	0.68	531	242	0.46	0.00
469	scaff31227 complement(1300..1453)	CHH	6	9	5	0.56	200	45	0.23	0.04
470	scaff33439 112313..112380	CHH	19	98	10	0.10	190	41	0.22	0.01
471	scaff30955 457..612	CHH	45	255	24	0.09	1602	323	0.20	0.00
472	scaff30955 83109..83165	CHH	18	183	0	0.00	107	4	0.04	0.02
473	scaff30578 complement(89923..90027)	CHH	8	74	9	0.12	497	115	0.23	0.02
474	scaff31840 complement(63043..63902)	CHH	6	41	2	0.05	94	27	0.29	0.00
475	scaff33998 complement(85281..85630)	CHH	10	45	1	0.02	5	2	0.40	0.02
476	scaff33850 complement(83239..83747)	CHH	142	317	163	0.51	2051	1193	0.58	0.01
477	scaff31982 complement(51574..51907)	CHH	77	597	145	0.24	2529	703	0.28	0.05
478	scaff31982 complement(54430..54546)	CHH	35	201	46	0.23	648	220	0.34	0.00
479	scaff33291 complement(42576..42900)	CHH	14	12	3	0.25	91	53	0.58	0.03
480	scaff31476 complement(36306..36973)	CHH	122	93	0	0.00	961	66	0.07	0.00
481	scaff39257 40386..40578	CHH	47	154	0	0.00	853	58	0.07	0.00
482	scaff39257 40386..40632	CHH	19	134	0	0.00	235	16	0.07	0.00
483	scaff37750 complement(19291..19873)	CHH	68	176	58	0.33	784	383	0.49	0.00
484	scaff37750 complement(19945..20114)	CHH	48	180	56	0.31	1350	519	0.38	0.03
485	scaff31978 37289..37357	CHH	20	119	57	0.48	477	282	0.59	0.02
486	scaff35973 complement(37502..37735)	CHH	38	74	0	0.00	265	12	0.05	0.05
487	scaff31977 11948..12305	CHH	112	542	226	0.42	2973	1398	0.47	0.01
488	scaff35919 10973..11045	CHH	15	49	8	0.16	277	95	0.34	0.01
489	scaff34669 complement(4760..4894)	CHH	35	195	25	0.13	302	65	0.22	0.01
490	scaff34669 4987..5038	CHH	15	132	20	0.15	561	151	0.27	0.00
491	scaff30795 complement(3614..3948)	CHH	97	78	0	0.00	712	123	0.17	0.00
492	scaff35161 4010..4371	CHH	63	49	18	0.37	747	456	0.61	0.00
493	scaff32804 complement(5598..5635)	CHH	14	19	1	0.05	315	105	0.33	0.01
494	scaff32804 complement(6086..6316)	CHH	28	12	0	0.00	569	165	0.29	0.02
495	scaff32804 complement(25319..25391)	CHH	20	32	6	0.19	290	106	0.37	0.03
496	scaff35218 24406..24719	CHH	41	201	1	0.00	590	38	0.06	0.00
497	scaff34503 10971..11025	CHH	19	130	21	0.16	748	228	0.30	0.00
498	scaff35979 complement(17659..17973)	CHH	69	46	18	0.39	919	487	0.53	0.05
499	scaff33842 complement(19929..20066)	CHH	35	6	0	0.00	370	172	0.46	0.02
500	scaff32090 complement(309..345)	CHH	9	63	14	0.22	374	139	0.37	0.01
501	scaff32090 complement(4354..4452)	CHH	30	194	16	0.08	876	197	0.22	0.00
502	scaff30556 17353..17511	CHH	28	106	47	0.44	325	181	0.56	0.03
503	scaff37199 complement(20955..21012)	CHH	20	74	21	0.28	366	151	0.41	0.02
504	scaff40518 3189..15987	CHH	2462	3925	1794	0.46	28948	14181	0.49	0.00
505	scaff40518 20062..20511	CHH	29	20	6	0.30	256	154	0.60	0.01
506	scaff38867 15177..15229	CHH	6	45	0	0.00	27	4	0.15	0.02
507	scaff31377 1110..1227	CHH	44	360	84	0.23	2389	769	0.32	0.00
508	scaff35064 complement(5890..6018)	CHH	50	1083	10	0.01	2933	120	0.04	0.00
509	scaff33434 1018..1079	CHH	17	68	0	0.00	564	99	0.18	0.00
510	scaff33989 complement(429..787)	CHH	73	167	19	0.11	840	357	0.43	0.00
511	scaff51495 5703..5815	CHH	41	153	2	0.01	280	19	0.07	0.01
512	scaff32014 551..592	CHH	11	58	5	0.09	268	73	0.27	0.00
513	scaff38081 complement(1492..1532)	CHH	9	96	32	0.33	307	146	0.48	0.01
514	scaff38081 complement(1492..1622)	CHH	29	173	43	0.25	671	222	0.33	0.02
515	scaff34277 complement(9665..10274)	CHH	63	146	7	0.05	792	98	0.12	0.00
516	scaff32988 9883..10022	CHH	43	422	9	0.02	1830	78	0.04	0.02
517	scaff32473 complement(6673..7279)	CHH	155	955	144	0.15	6136	1081	0.18	0.03
518	scaff32473 complement(8360..8437)	CHH	25	194	1	0.01	629	57	0.09	0.00
519	scaff37939 complement(471..536)	CHH	20	162	30	0.19	1371	402	0.29	0.00
520	scaff37939 2543..3234	CHH	142	607	242	0.40	2754	1228	0.45	0.02

521	scaff41091 2912..2985	CHH	20	88	15	0.17	748	207	0.28	0.02
522	scaff41091 2912..3003	CHH	6	18	0	0.00	246	54	0.22	0.01
523	scaff41091 complement(4672..4725)	CHH	18	226	59	0.26	850	276	0.32	0.04
524	scaff34713 2794..3122	CHH	49	63	13	0.21	566	218	0.39	0.00
525	scaff32877 6493..6556	CHH	12	136	27	0.20	302	85	0.28	0.04
526	scaff35068 3238..3675	CHH	64	31	6	0.19	503	245	0.49	0.00
527	scaff35039 4616..4798	CHH	47	125	11	0.09	766	167	0.22	0.00
528	scaff31824 complement(696..799)	CHH	29	137	31	0.23	424	138	0.33	0.02
529	scaff31611 7539..7577	CHH	13	149	0	0.00	512	12	0.02	0.05
530	scaff40088 3923..4079	CHH	24	90	0	0.00	291	20	0.07	0.00
531	scaff30487 complement(2314..2399)	CHH	27	445	44	0.10	1642	224	0.14	0.02
532	scaff42114 8632..9475	CHH	195	30	9	0.30	1524	824	0.54	0.01
533	scaff39285 824..863	CHH	11	11	0	0.00	129	75	0.58	0.00
534	scaff43474 complement(5950..6160)	CHH	67	689	303	0.44	2704	1292	0.48	0.04
535	scaff44791 5940..6228	CHH	70	176	13	0.07	1585	212	0.13	0.01
536	scaff32202 complement(6614..6994)	CHH	40	71	0	0.00	602	119	0.20	0.00
537	scaff32886 complement(2040..2108)	CHH	20	85	0	0.00	134	8	0.06	0.02
538	scaff45248 3362..3468	CHH	33	327	95	0.29	476	176	0.37	0.01
539	scaff43229 1768..1906	CHH	23	298	45	0.15	1117	231	0.21	0.02
540	scaff35746 5262..5316	CHH	20	83	9	0.11	321	104	0.32	0.00
541	scaff31751 4834..5063	CHH	43	195	6	0.03	689	48	0.07	0.03
542	scaff50495 5530..5844	CHH	13	25	0	0.00	56	8	0.14	0.04
543	scaff52787 2950..3234	CHH	96	163	25	0.15	1961	462	0.24	0.01
544	scaff37164 3153..3931	CHH	125	87	12	0.14	911	499	0.55	0.00
545	scaff30930 complement(4761..4933)	CHH	50	446	17	0.04	2175	139	0.06	0.02
546	scaff52028 complement(2777..2855)	CHH	27	22	1	0.05	483	161	0.33	0.00
547	scaff35638 complement(2787..2854)	CHH	16	98	7	0.07	626	88	0.14	0.04
548	scaff30845 complement(1258..1339)	CHH	21	40	0	0.00	467	97	0.21	0.00
549	scaff32780 869..955	CHH	27	15	3	0.20	257	117	0.46	0.04
550	scaff34748 complement(2049..3836)	CHH	330	1302	77	0.06	8711	1245	0.14	0.00
551	scaff46875 complement(119..270)	CHH	38	223	3	0.01	603	23	0.04	0.05
552	scaff37887 602..1246	CHH	52	29	2	0.07	593	168	0.28	0.01
553	scaff42657 1014..1059	CHH	6	47	3	0.06	262	57	0.22	0.01
554	scaff52755 complement(1891..1995)	CHH	39	51	2	0.04	550	119	0.22	0.00
555	scaff38770 230..373	CHH	13	62	3	0.05	85	15	0.18	0.02
556	scaff38770 1925..1994	CHH	17	92	11	0.12	536	143	0.27	0.00
557	scaff35201 1397..1464	CHH	21	83	30	0.36	919	446	0.49	0.02
558	scaff47196 1423..1474	CHH	14	115	25	0.22	505	159	0.31	0.02
559	scaff52783 746..861	CHH	33	122	14	0.11	1150	285	0.25	0.00
560	scaff52783 1087..1139	CHH	8	98	13	0.13	362	81	0.22	0.03
561	scaff47024 800..957	CHH	52	234	79	0.34	1635	745	0.46	0.00
562	scaff36976 complement(1822..1908)	CHH	21	15	5	0.33	143	98	0.69	0.01
563	scaff43394 311..719	CHH	123	722	90	0.12	3743	564	0.15	0.04
564	scaff36153 complement(1659..1699)	CHH	17	30	3	0.10	165	49	0.30	0.02
565	scaff37071 937..981	CHH	14	32	2	0.06	147	34	0.23	0.02
566	scaff39354 623..1625	CHH	293	1622	21	0.01	10188	218	0.02	0.01
567	scaff45022 complement(712..1432)	CHH	154	1483	26	0.02	890	30	0.03	0.01
568	scaff45207 complement(352..414)	CHH	21	15	3	0.20	219	106	0.48	0.03
569	scaff41617 complement(977..1029)	CHH	9	44	2	0.05	282	89	0.32	0.00
570	scaff51220 833..879	CHH	14	411	127	0.31	1936	718	0.37	0.01
571	scaff37112 862..916	CHH	19	166	34	0.20	761	240	0.32	0.00
572	scaff47175 766..957	CHH	37	201	42	0.21	975	279	0.29	0.01
573	scaff47331 complement(1573..1609)	CHH	6	189	59	0.31	729	308	0.42	0.00
574	scaff47331 complement(1573..1639)	CHH	8	279	118	0.42	1173	563	0.48	0.05
575	scaff52654 complement(644..722)	CHH	19	149	31	0.21	722	213	0.30	0.02
576	scaff31810 1499..1550	CHH	10	107	5	0.05	265	29	0.11	0.04
577	scaff51115 204..318	CHH	34	45	0	0.00	375	88	0.23	0.00
578	scaff32397 complement(722..792)	CHH	17	41	5	0.12	450	114	0.25	0.04
579	scaff51437 1007..1104	CHH	32	64	9	0.14	522	167	0.32	0.00
580	scaff36799 19..355	CHH	36	40	0	0.00	352	50	0.14	0.00
581	scaff38205 1458..1518	CHH	22	372	90	0.24	1696	625	0.37	0.00
582	scaff36173 695..819	CHH	27	106	48	0.45	474	281	0.59	0.01
583	scaff48247 complement(451..506)	CHH	10	25	2	0.08	67	18	0.27	0.04
584	scaff52061 complement(1578..1852)	CHH	99	1083	112	0.10	6396	919	0.14	0.00
585	scaff48398 complement(228..292)	CHH	18	55	11	0.20	512	184	0.36	0.01
586	scaff48398 698..786	CHH	27	324	87	0.27	1536	625	0.41	0.00
587	scaff41332 complement(647..716)	CHH	25	36	0	0.00	416	45	0.11	0.02
588	scaff49426 complement(556..845)	CHH	86	658	172	0.26	2883	1005	0.35	0.00
589	scaff45514 568..655	CHH	26	205	24	0.12	1233	298	0.24	0.00
590	scaff45514 1069..1110	CHH	17	298	49	0.16	1044	233	0.22	0.02
591	scaff49645 complement(1036..1096)	CHH	23	170	27	0.16	880	192	0.22	0.05
592	scaff44271 187..333	CHH	34	369	4	0.01	195	8	0.04	0.02
593	scaff31227 complement(793..943)	CHH	48	43	1	0.02	1148	378	0.33	0.00

Table S39:TE with expression and methylation level.

Gene	Description	A	TB	R	TB	R
		Expression level (log ₂ RPM)			Methylation level	
g2381	T18347gag homolog, gypsy retrotransposon	0.65	-1.5	-2.6	0.15	0.40
g2744	Transposon Tf2-1 poly	2.87	4.55	3.89	0.18	0.00
g2998	Transposon, putative	1.78	1.68	1.46	0.20	0.32
g6485	Retrovirus-related Pol poly from transposon TNT 1-94	-2.67	-3.3	-2.3	0.54	0.00
g7277	Transposon Tf2-1 poly	4.57	4.56	4.43	0.14	0.00
g8973	retrotransposon nucleocapsid	4.52	5.61	5.32	0.24	0.51
g8976	Retrovirus-related Pol poly from transposon TNT 1-94	0.59	-0.1	2.02	0.22	0.50
g9058	retrovirus-related pol poly from transposon tnt 1-94	-1.75	-2.5	-2.5	0.18	0.26
g10356	Retrovirus-related Pol poly from transposon TNT 1-94	3.22	2.82	2.47	0.20	0.00
g10359	Retrovirus-related Pol poly from transposon TNT 1-94	2.89	2.75	2.62	0.16	0.00
g10489	Retrovirus-related Pol poly from transposon TNT 1-94	3.16	-1.7	-0.7	0.17	0.22
g11240	Transposon, putative	-0.99	-1.3	-1.6	0.13	0.00
g12818	Retrovirus-related Pol poly from transposon TNT 1-94	-0.27	2.21	1.59	0.15	0.25
g12819	Retrovirus-related Pol poly from transposon TNT 1-94	2.16	3.66	3.15	0.14	0.00
g13579	Retrovirus-related Pol poly from transposon TNT 1-94	-1.68	-1.4	-1.8	0.42	0.42
g14070	retrotransposon nucleocapsid	2.62	-0.8	1.8	0.65	0.72
g14370	Retrovirus-related Pol poly from transposon TNT 1-94	3.58	3.6	3.67	0.15	0.00
g14565	retrovirus-related pol poly from transposon tnt 1-94	2	2.41	1.54	0.12	0.16
g14891	Retrovirus-related Pol poly from transposon TNT 1-94	1.74	-0.1	-0.2	0.23	0.31
g14914	Retrovirus-related Pol poly from transposon TNT 1-94	0.21	-0.3	-0.7	0.17	0.17
g14982	retrovirus-related pol poly from transposon tnt 1-94	3.7	3.19	2.53	0.19	0.21
g15324	retrovirus-related Pol poly from transposon TNT 1-94	0.72	1.38	1.77	0.21	0.23
g15763	Retrovirus-related Pol poly from transposon TNT 1-94	3.45	1.14	0.84	0.11	0.16

Methylation level=methylated coverage/context coverage. For each TE, methylation level in thallus body (TB) and rhizine (R) tissues and expression levels in apothecia (A), thallus body (TB) and rhizine (R) tissues are provided.

Table S40: List of all the transporters in *P. membranacea* and their expression in A, TB and R tissues.

Gene id	Description	A (log2 RPM)	TB (log2 RPM)	R (log2 RPM)
g2902	ABC multidrug transporter	0.28	3.67	3.3
g3989	ABC multidrug transporter	5.67	5.41	5.67
g581	ABC transporter	7.04	2.77	2.69
g1052	ABC transporter	5.58	5.35	1.71
g1391	ABC transporter	2.84	2.3	3.06
g10653	ABC transporter	6.75	6.44	6.34
g11974	ABC transporter	5.4	4.97	4.75
g9538	ABC transporter (Adp1)	6.87	6.32	6.51
g236	ABC transporter ATP-binding ARB1	7.26	6.87	6.82
g11645	ABC transporter F family member 3	5.59	4.72	4.62
g8478	ABC transporter family	5.31	5.64	5.5
g11788	ABC transporter G family member 11	3.62	2.44	2.84
g4341	ABC-transporter	4.42	5.01	5.56
g5557	ABC-transporter	4.72	5.1	5.04
g13400	ABC-transporter	-1.08	1.79	1.9
g961	Amino acid polyamine transporter I	6.07	6.15	5.99
g4050	amino acid transporter	4.2	5.62	4.89

g11232	amino acid transporter	5.84	5.92	5.9
g6837	amino acid transporter amino acid transport and metabolism	4.95	5.29	5.35
g10257	calcium permease family membrane transporter	6.95	6.46	6.52
g4404	calcium transporter,	5.59	4.15	2.69
g3031	cation chloride cotransporter	5.77	5.3	5.28
g6422	cation transporter	3.27	3.34	3.33
g11754	Chromate transporter	1.96	2.38	2.42
g4522	Ctr copper transporter	6.66	6.72	6.93
g6731	Ctr copper transporter	6.49	6.9	7.54
g3107	Ctr copper transporter family	6.84	6.77	7.46
g14156	Dityrosine transporter 1	4.89	4.95	4.01
g12507	Drug metabolite transporter	5.38	5.54	5.71
g9750	family transporter	5.88	7.7	7.79
g13250	Fe2 Zn2 regulated transporter	7.86	6.19	6.36
g9473	FLVCR family MFS transporter	4.65	5.04	5
g1090	GDP-mannose transporter GONST5 Q9SFE9	3.78	1.55	1.53
g527	general substrate transporter	4.51	4.43	4.12
g3985	general substrate transporter	5.42	6.96	5.64
g4882	general substrate transporter	1.82	-1.87	-1.85
g4914	General substrate transporter	2.17	-0.26	-0.93
g6341	general substrate transporter	6.81	7.56	7.91
g12121	general substrate transporter	3.52	-2.29	-1.4
g14395	general substrate transporter	5.12	5.22	4.89

g15965	general substrate transporter	3.52	3.85	3.98
g16043	general substrate transporter	1.63	-0.95	-0.93
g12973	hexose transporter of the major facilitator superfamily	1.85	-0.73	-0.71
g3059	high-affinity iron transporter	5.48	6.11	6.03
g917	inner membrane magnesium transporter mrs2	5.52	7.01	6.94
g950	inner membrane magnesium transporter mrs2	5.45	4.81	5.12
g490	Inorganic phosphate transporter PHO84	5.65	5.3	5.37
g981	intracellular protein transporter	6.09	5.73	5.64
g2534	iron-sulfur clusters transporter atm1	5.29	5.51	5.07
g6520	lactate pyruvate transporter	6.85	6.8	6.96
g1641	Long-chain fatty acid transporter	6.15	5.61	5.83
g6344	lysosomal cobalamin transporter	7.75	7.69	7.97
g4274	Major facilitator superfamily domain general substrate transporter	5.07	4.79	4.77
g12410	Major facilitator superfamily domain, general substrate transporter	1.98	3.36	1.53
g4335	Major facilitator superfamily general substrate transporter	6.37	6.22	5.69
g3782	major facilitator superfamily transporter	6.48	6.45	6.19
g4924	major facilitator superfamily transporter	4.59	3.38	3.21
g6018	major facilitator superfamily transporter	5.53	6.92	4.52
g13735	Major facilitator superfamily, general substrate transporter domain-containing	3.35	2.88	4.98
g9751	mate family transporter	4.99	7.27	7.38
g7427	membrane transporter	6.31	6.6	6.49
g6379	metabolite transporter	5.93	2.36	1.19

g7210	metal ion transporter	4.98	5.75	5.73
g3974	metal ion transporter metal ion transporter	6.25	5.6	5.62
g201	MFS general substrate transporter	6.84	7.6	7.81
g390	MFS general substrate transporter	5.08	4.6	4.59
g697	MFS general substrate transporter	6.58	5.58	5.48
g829	MFS general substrate transporter	6.91	6.68	6.41
g849	MFS general substrate transporter	5.89	5.36	5.46
g1056	MFS general substrate transporter	5.14	5.37	4.34
g1184	MFS general substrate transporter	6.57	3.19	3.67
g1738	MFS general substrate transporter	4.55	6.19	6.08
g2378	MFS general substrate transporter	5.36	2.4	1.25
g2379	MFS general substrate transporter	5.03	2.08	1.12
g2677	MFS general substrate transporter	4.97	5.71	6.36
g3100	MFS general substrate transporter	5.46	6.34	6.21
g3101	MFS general substrate transporter	5.08	5.5	5.47
g3539	MFS general substrate transporter	4.16	0.46	2.29
g4289	MFS general substrate transporter	7.21	7.32	7.5
g4529	MFS general substrate transporter	3.7	5.67	6.6
g4530	MFS general substrate transporter	4.8	5.81	6.9
g5025	MFS general substrate transporter	5.03	5.15	5.35
g5433	MFS general substrate transporter	6.93	5.87	5.47
g5951	MFS general substrate transporter	7.04	6.84	6.89
g6027	mfs general substrate transporter	2.8	0.67	1.92

g6183	MFS general substrate transporter	5.93	6.83	6.28
g6291	MFS general substrate transporter	3.35	4.72	4.13
g7123	MFS general substrate transporter	8.32	4.82	5.99
g7473	MFS general substrate transporter	6.22	5.75	6.37
g7825	MFS general substrate transporter	3.55	4.62	4.89
g7826	MFS general substrate transporter	4.08	4.91	5.13
g9223	MFS general substrate transporter	4.05	4.02	4.33
g9345	MFS general substrate transporter	5.54	5.87	4.37
g10102	MFS general substrate transporter	6.95	6.8	6.94
g10745	MFS general substrate transporter	6.45	6.69	6.84
g10853	MFS general substrate transporter	6.09	5.48	4.31
g11304	MFS general substrate transporter	3.02	3.14	2.48
g11305	MFS general substrate transporter	2.91	3.27	2.89
g11313	MFS general substrate transporter	6.39	6.88	6.98
g11690	MFS general substrate transporter	6.27	6.01	5.98
g13000	MFS general substrate transporter	4.69	5.22	4.98
g13421	MFS general substrate transporter	5.14	5.76	6.39
g13592	MFS general substrate transporter	3.7	4.36	3.01
g13856	MFS general substrate transporter	6.07	6.41	6.84
g13857	MFS general substrate transporter	6.07	6.54	6.89
g13986	MFS general substrate transporter	2.62	1.79	3.41
g14078	MFS general substrate transporter	5.4	5.05	5.8
g14582	MFS general substrate transporter	2.7	3.41	3.14

g14840	MFS general substrate transporter	5.68	5.15	4.63
g9566	mfs monocarboxylate transporter	3.76	3.97	3.78
g1662	MFS monocarboxylic acid transporter	5.71	5.48	5.78
g11259	MFS multidrug resistance transporter,	4.44	5.36	7.13
g2602	mfs multidrug transporter	3.55	4.38	4.97
g6340	MFS multidrug transporter	5.26	4.52	5.35
g8273	MFS multidrug transporter	5.87	5.8	5.91
g9583	MFS multidrug transporter	6.53	6.49	6.24
g12349	MFS multidrug transporter	5.46	5.73	6.3
g13900	MFS multidrug transporter	5.08	4.09	4.48
g15782	MFS multidrug transporter	2.37	3.06	3.53
g16282	mfs multidrug transporter	6.61	7.08	4.56
g2603	MFS multidrug transporter,	4.38	4.49	4.99
g5990	MFS phospholipid transporter	6.11	5.5	5.03
g1872	MFS transporter	5.81	5.1	5.11
g4312	mfs transporter	6.07	5.48	5.83
g4313	MFS transporter	6.02	4.95	5.34
g5007	MFS transporter	5.42	5.39	5.24
g5653	MFS transporter	0	-0.64	-0.62
g6765	MFS transporter	7.76	7.51	7.38
g7307	MFS transporter	5.67	6.03	6.17
g7870	MFS transporter	5.25	6.07	6.08
g8872	MFS transporter	5.28	6.1	6.34

g9272	MFS transporter	-0.21	3.28	2.04
g13377	MFS transporter	3.87	4.24	0.01
g15995	MFS transporter	3.43	6.05	2.79
g10545	MFS transporter,	6.43	4.12	3.8
g11529	MFS transporter,	1.91	2.81	1.34
g11771	MFS transporter,	5.44	5.65	5.59
g11954	MFS transporter, ACS family, allantoate permease	3.39	6.12	2.86
g8423	MFS transporter, SIT family, siderophore-iron:H ⁺ symporter	0.52	-0.77	-0.75
g10323	Molybdate transporter	7.07	7.39	7.48
g10974	Monocarboxylate transporter 10	6.13	6.48	6.46
g3927	monocarboxylate transporter,	2.7	2.98	1.58
g9187	multidrug efflux transporter transmembrane domain-containing	6.1	6.15	6.29
g241	multidrug transporter	3.51	0.8	3.01
g4757	nitrate transporter	4.9	5.73	5.98
g405	Nonglycosylated anion transporter	3.92	4.36	4.04
g2005	nucleoside transporter family	5.22	5.68	5.67
g1679	nucleotide-sugar transporter	5.46	4.57	4.8
g3529	oligopeptide transporter	4.28	2.82	2.91
g8932	oligopeptide transporter	0.8	-0.5	-0.47
g4594	OPT family small oligopeptide transporter	5.89	6.63	6.71
g6190	outer membrane autotransporter	5.45	5.75	6.87
g2638	outer membrane autotransporter barrel	2.47	3.04	3.08
g8218	outer membrane autotransporter barrel	8.56	8.67	6.5

g297	pantothenate transporter liz1	4.34	4.35	4.34
g6767	pantothenate transporter liz1	0.12	0.87	1.48
g9188	patched sphingolipid transporter	6.09	5.82	5.8
g9189	patched sphingolipid transporter (Ncr1),	6.27	6.26	6.31
g13142	Peptide transporter PTR2 [Candida albicans]	4.66	3.62	3.75
g13714	phospholipid P-type ATPase transporter	5	4.17	4.3
g9061	phthalate transporter	6.01	3.35	2.34
g12987	Pi-transporter A-1	7.23	7.83	8.18
g10208	plasma membrane zinc ion transporter,	4.1	5.26	4.99
g5552	probable major myo-inositol transporter iolT	6.46	6.28	6.9
g1744	putative transporter	4.9	6.24	6.36
g7159	putative transporter	-0.13	0.59	1.49
g13955	putative transporter C3H1.06c	2.67	-0.49	-1.05
g1169	related to 4-hydroxybenzoate transporter	6.78	6.49	6.61
g5556	related to ABC transporter	4.5	4.66	4.36
g12902	related to dityrosine transporter	5.44	5.29	6
g9041	related to major facilitator (MFS1) transporter	5.74	5.65	5.18
g8344	related to MFS transporter	6.41	6.3	6.59
g1280	related to Mg2 transporter -like	4.75	5.51	5.24
g523	retrograde transporter	6.29	5.94	6.4
g113	Saccharomyces cerevisiae YBR180W DTR1 dityrosine transporter	3.89	5.67	5.24
g463	siderophore iron transporter mirC	1.63	0.58	-0.65
g4256	sodium calcium transporter	6.1	6.98	6.7

g48	sodium nucleoside cotransporter 1 nucleoside cotransporter 1	5.02	5.99	5.17
g679	sphingoid long-chain base transporter RSB1	4.64	6.25	6.63
g8557	sphingoid long-chain base transporter RSB1	4.76	3.97	3.98
g4825	succinate fumarate mitochondrial transporter	4.43	3.42	3.09
g6528	sulfate transporter	5.46	5.5	5.88
g6529	sulfate transporter	5.76	5.17	5.46
g11342	sulfate transporter family	5.18	4.59	4.49
g9039	tetracycline-efflux transporter	6.44	6.27	6.21
g9626	transporter C460,05	5.21	6.21	5.9
g11216	transporter SEC23	6.03	5.78	5.96
g1952	transporter SEC31	5.93	6.13	6.24
g13555	transporter YIL166C	3.43	3.51	2.45
g417	triose-phosphate transporter	6.28	6.73	6.9
g9892	triose-phosphate transporter	6.88	6.59	6.8
g1814	UAA transporter	4.43	5.82	6.11
g1962	UAA transporter, partial	3.64	4.54	4.63
g12041	UDP-galactose transporter	7.21	45.41	23.32
g12356	UDP-galactose transporter	6.77	7.13	7.12
g14788	UDP-N-acetylglucosamine transporter YEA4	2.01	-2.16	-2.14
g5155	Uncharacterized transporter	3.92	5.86	2.41
g13036	urea transporter	5.09	6.59	6.5
g5208	vacuolar amino acid transporter 1	5.36	5.14	4.86
g9416	vacuolar transporter chaperone	5.59	5.32	5.74

g8718	vesicle transporter SEC22	6.91	6.62	6.63
g13416	zinc ion transporter Zhf1	1.34	1.03	1.31
g6343	zinc transporter	5.2	4.73	5.15
g6482	zip zinc transporter	6.18	6.31	6.35

Appendix L: KEGG maps.

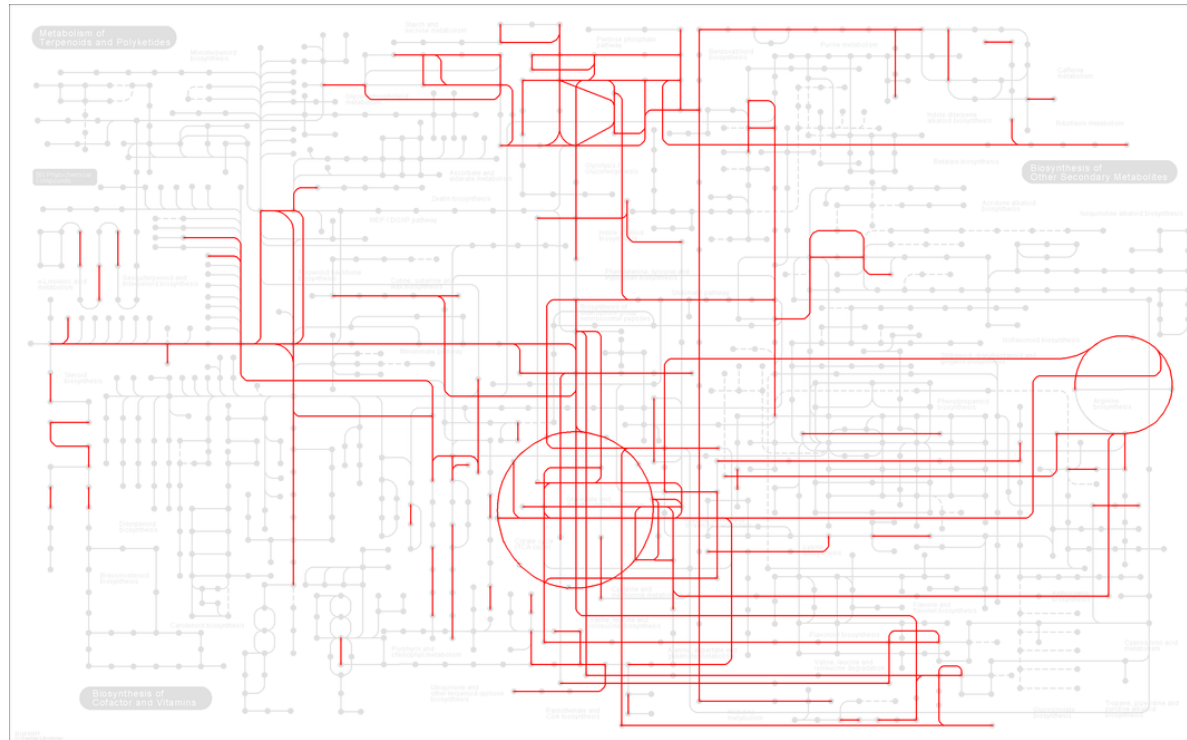


Figure S24: KEGG map of Biosynthesis of secondary metabolites identified in *Peltigera membranacea*. In black, the KO reference pathway and in red, the pathway represented in *Peltigera membranacea*.

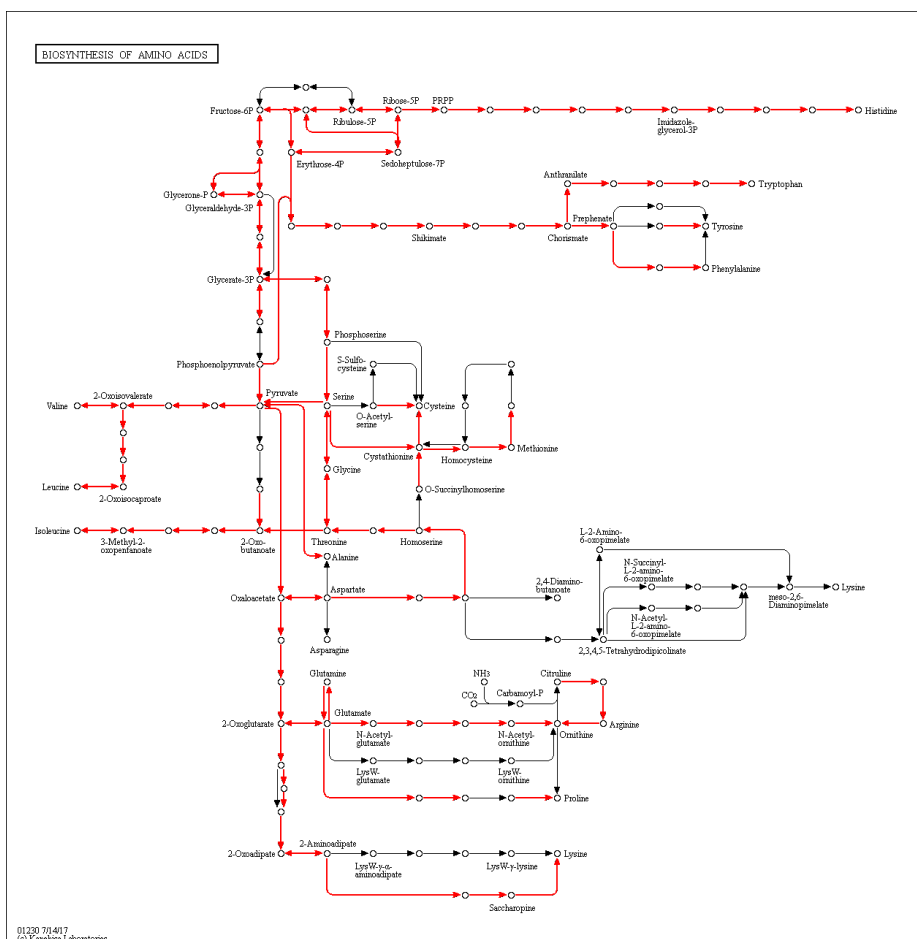


Figure S25: KEGG map of Biosynthesis of amino acids identified in *Peltigera membranacea*. In black, the KO reference pathway and in red, the pathway represented in *Peltigera membranacea*.

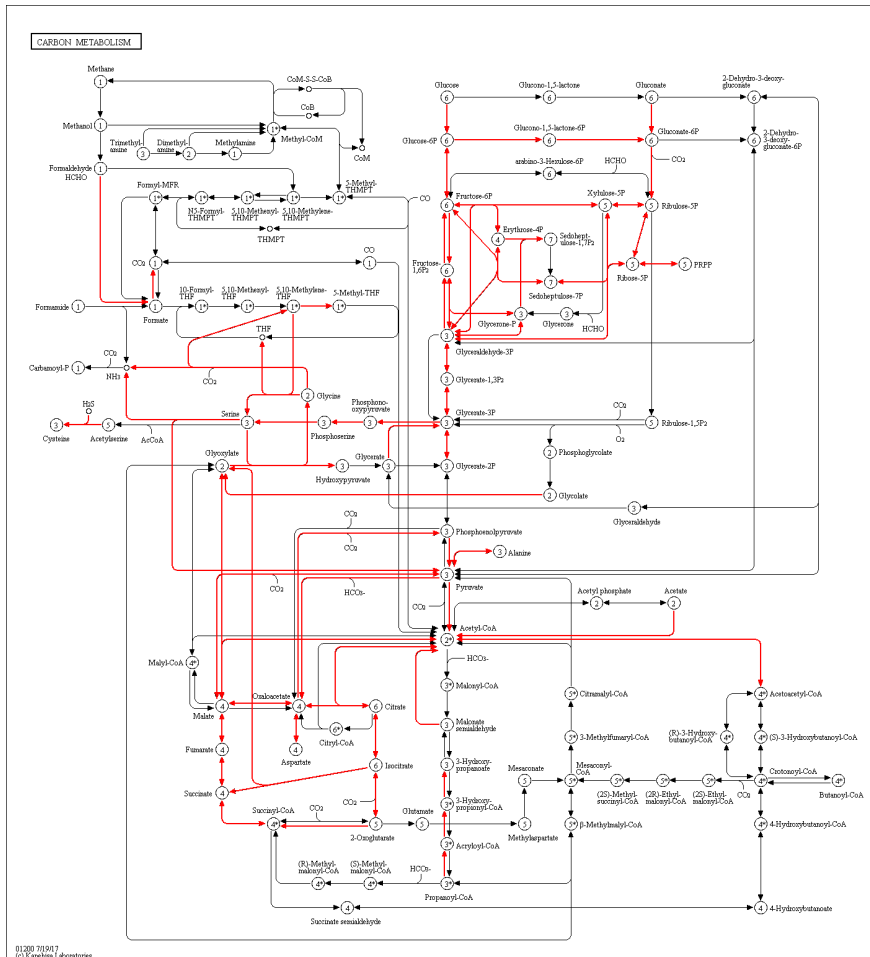


Figure S27: KEGG map of Carbon metabolism identified in *Peltigera membranacea*.
In black, the KO reference pathway and in red, the pathway represented in *Peltigera membranacea*.

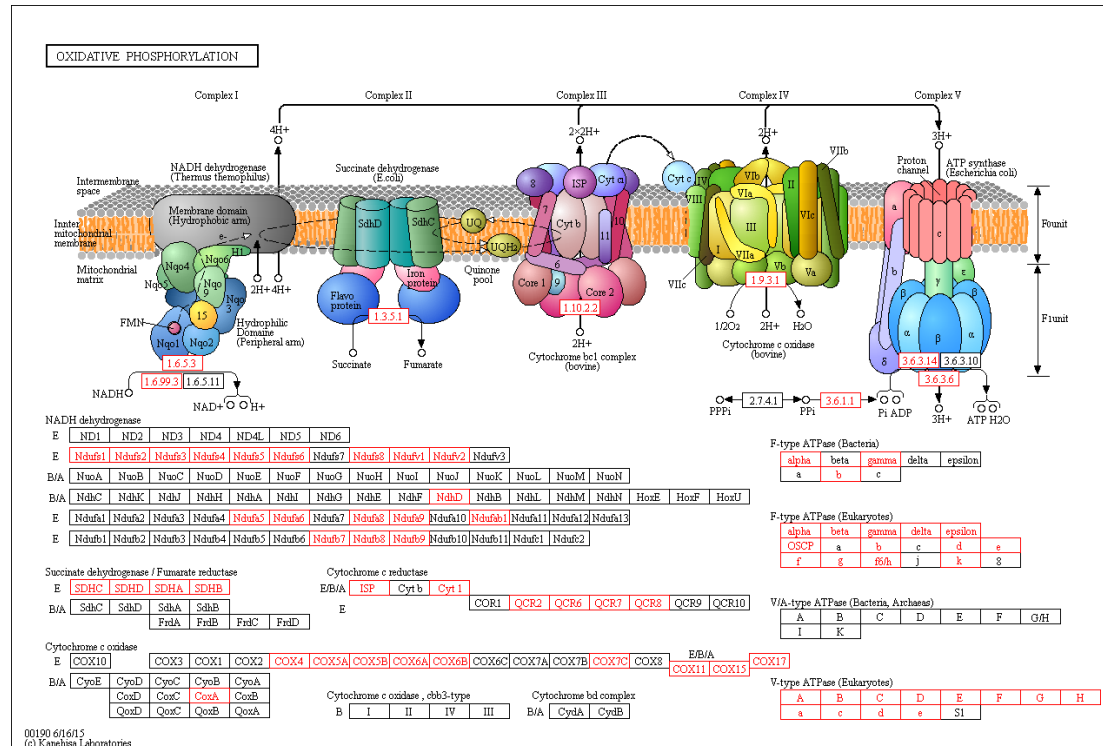


Figure S28: KEGG map of Oxidative phosphorylation identified in *Peltigera membranacea*.
In black, the KO reference pathway and in red, the pathway represented in *Peltigera membranacea*.

Appendix M: blast2go2gff.pl

```
1  #!/usr/bin/perl
2  use warnings;
3  use strict;
4
5  my ($file) = @ARGV;
6
7  unless ($file) {
8      die "Usage is blast2go2gff.pl [blast2go text file]\n";
9  }
10
11 my $outfile = "$file.gff";
12
13
14 open (IN,$file) or die "Can't open $file: $!";
15
16 open (OUT,'>',$outfile) or die "Can't write to $outfile: $!";
17
18 while (<IN>) {
19
20     chomp;
21     s/[\r\n]//g;
22
23     my ($name,$desc) = split(/\t/);
24
25     # We need to split the name into scaffold name, strand and positions
26
27     my $strand;
28     my $chr;
29     my $positions;
30     my $cds_name;
31     if ($name =~ /^(g\d+)(.*)_CDS_join\(((\d\.|,|+))\)/) {
32         $strand = '+';
33         $cds_name = $1;
34         $chr = $2;
35         $positions = $3;
36     }
37
38     elsif ($name =~ /^(g\d+)(.*)_CDS_complement\((\join\(((\d\.|,|+))\))\)/) {
39         $strand = '-';
40         $cds_name = $1;
41         $chr = $2;
42         $positions = $3;
43     }
44
45     else {
46         warn "Couldn't get position from $name\n";
47         next;
48     }
49
50     # Now we need to separate the positions
51     my @positions = split(/,/,$positions);
52
53     foreach my $position(@positions) {
54         my ($start,$end) = split(/\./,$position);
55
56         # print out a suitable GFF line.
57
58         print OUT join("\t",($chr,'blast2go2gff','CDS',$start,$end,'.',
59             $strand,'.','transcript_id=$cds_name; description=$desc; name=$cds_name')), "\n";
60
61     }
62
63 }
```

Table S41: List of PKSs in *P. membranacea* (P.mem v1.1) genome.

Geneid	Description	Apothecia	Thallus Body	Rhizine
g4511	non-reducing type I polyketide synthase	5.71	6.48	8.63
g6020	reducing type I polyketide synthase	3.57	3.70	0.96
g6021	reducing type I polyketide synthase	3.62	4.07	0.50
g6022	reducing type I polyketide synthase	4.05	4.52	1.63
g6716	non-reducing type I polyketide synthase	2.04	-1.21	-1.19
g7937	reducing type I polyketide synthase	3.95	5.16	5.37
g8166	reducing type I polyketide synthase	1.26	2.30	1.67
g8167	reducing type I polyketide synthase	1.35	2.53	2.11
g8168	reducing type I polyketide synthase	1.91	3.49	2.91
g8169	reducing type I polyketide synthase	1.66	3.75	3.39
g8391	reducing type I polyketide synthase	-1.39	-6.75	-6.73
g10306	reducing type I polyketide synthase 10	1.39	0.40	-2.32
g10421	reducing type I polyketide synthase	-0.16	0.64	0.68
g10449	reducing type I polyketide synthase 11	4.37	4.73	4.55
g10571	non-reducing type I polyketide synthase	2.06	2.80	0.85
g10700	reducing type I polyketide synthase	4.49	4.83	4.96
g10701	reducing type I polyketide synthase	5.18	5.71	5.69
g10702	reducing type I polyketide synthase	5.67	6.24	6.37
g12782	reducing type I polyketide synthase	0.38	1.46	1.40
g12783	reducing type I polyketide synthase	-0.36	1.49	1.10
g14729	reducing type I polyketide synthase	2.91	3.72	2.26
g14730	reducing type I polyketide synthase	2.64	3.33	2.67
g15202	reducing type I polyketide synthase	3.51	4.40	3.70
g15467	reducing type I polyketide synthase	3.58	4.77	4.01
g15875	reducing type I polyketide synthase	3.74	4.28	3.46