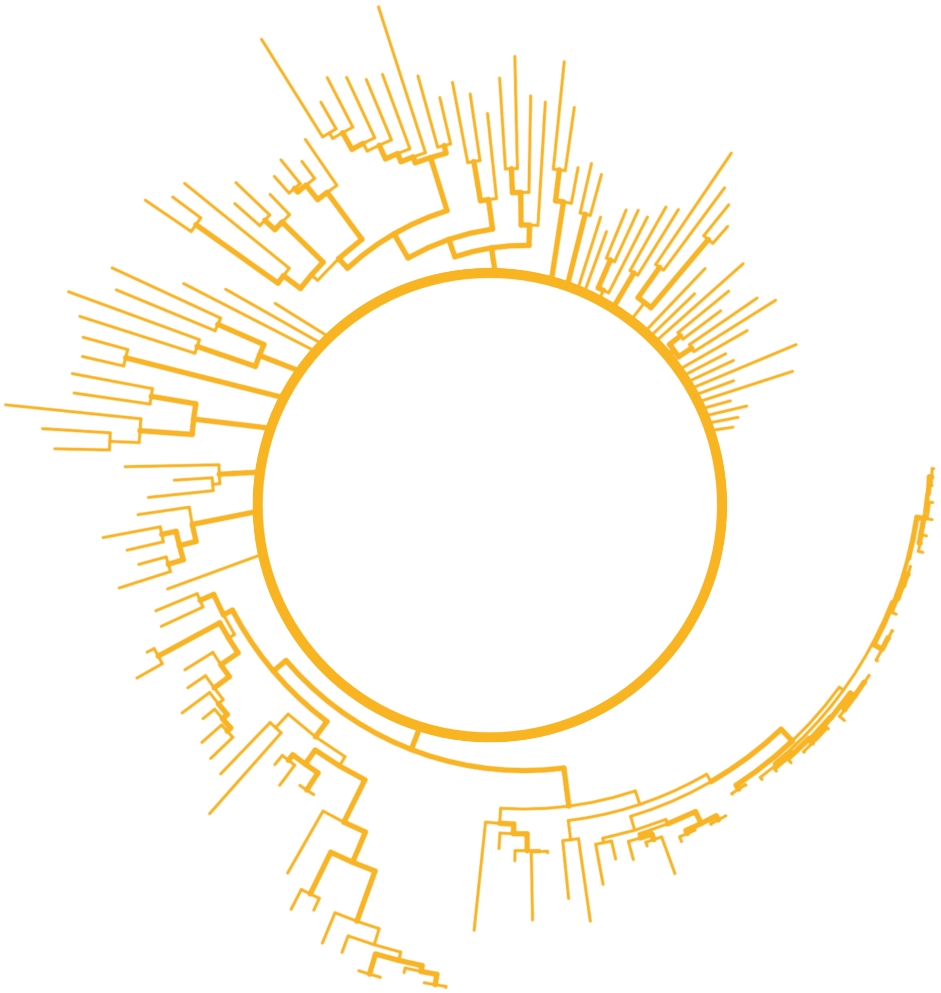
The Evolution of Anthraquinones as an Adaptive Trait in Lichen-forming Fungi



Theo Llewellyn

2023

Department of Life Sciences, Imperial College of Science, Technology and Medicine
Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and the Diploma of Imperial College London

Abstract

Fungi produce a remarkable diversity of chemical compounds to interact with their environment. Lichens—obligate mutualisms comprising fungi, photosynthetic organisms and bacteria—produce a particularly wide range of secondary compounds with functions including UV protection, anti-herbivory, antimicrobial activity, and metal homeostasis. Despite this chemical diversity, the genetic basis and evolutionary processes underpinning most lichen secondary metabolites have not been investigated. In this thesis, I focus on photoprotective anthraquinone pigments in the diverse *Teloschistales* order (*Ascomycota*) as a case study to explore how adaptive metabolic traits arise and then diversify in lichen-forming fungi.

Characterising the genes responsible for anthraquinone biosynthesis first required generating genome-scale data for the *Teloschistales*. Given the challenges of culturing lichen-forming fungi, I developed and implemented a lichen-specific metagenomics pipeline to sequence, assemble and annotate 24 new lichen-forming fungal genomes. Comparative genomics identified putative anthraquinone biosynthetic gene clusters (BGCs) in *Teloschistales* genomes and demonstrated that BGC diversification occurred via reshuffling existing enzyme genes with novel accessory genes. To understand anthraquinone evolution across the whole clade, I then expanded my metagenomic approach to sequence all major *Teloschistales* lineages. I combined this genomic dataset with densely sampled multilocus data to produce a robust genome-scale time tree. Phylogenomic analysis showed around half of current *Teloschistaceae* genera are not supported, and I propose a set of stable, evolutionarily relevant higher taxa instead. To understand how genomic variation affects the metabolite phenotype, I jointly analysed the genomes with new untargeted metabolome data. This revealed a complex interplay between genomic and metabolic variation and suggested that, for anthraquinones, BGC variation affects compound regulation and transport more than structural diversity. Finally, as anthraquinones are broadly cytotoxic, I hypothesised that anthraquinone-producing *Teloschistaceae* lichens evolved resistance mechanisms to avoid self-toxicity. Combining enzyme assays, axenic culture experiments, selection analysis and *in silico* protein modelling indicated that *Teloschistaceae* lichens achieved self-resistance through the evolution of efflux pumps, toxin methylation and resistant target enzymes. Together, my results demonstrate the power of multi-omic approaches to investigate the evolutionary processes that shape metabolite diversification in lichens.

Acknowledgments

Abans de tot, vull mostar el meu agraïment a l'Ester Gaya. Des del primer moment que ens vam conèixer vaig sentir la seva passió, bondat i amor per la ciència i els líquens. Estic segur que aquesta experiència es quedarà al meu cor per tota la vida. També vull donar les gràcies a la Facultat de Biologia de la UB, especialment als professors de Micologia, per obrir-me els ulls al món dels fongs. I would like to thank Tim Barraclough for his kindness, insightfulness and encouragement, and for making me feel like a true member of Barralab despite being based at a different institute. I would like to thank all the past and present members of Barralab, in particular Lily Peck, for her companionship and Reuben Nowell for his thoughtful guidance. I wish to thank Tom Prescott for agreeing to enter into the complicated world of lichens with us and for showing me a new perspective on science and the natural world. The Royal Botanic Gardens Kew has been central to my research career thus far, and I feel fortunate to have worked alongside many wonderful colleagues and plants. An extra thanks goes to Rowena Hill for being a friend. Thank you to all the tutors at the Evomics Workshop on Genomics and the Computational Molecular Evolution Workshop, especially Sandra Alvarez-Carretero, for guiding me so patiently through the world of Bayesian phylogenetics. A special thanks to the field lichenologists across the world who sent me samples after the pandemic almost derailed the whole project. Thank you to my friends and family for accompanying me on this journey and to Domani for being a continual source of joy.

Thank you to Jordi, Pascal and Simone per omplir la meva vie d'amour, luce e gioa.

Statement of Originality

I, Theo Llewellyn, confirm that the research included in this thesis is my own work. I have duly acknowledged any contributions carried out in collaboration with others. I have ensured that the work is original and any published material is clearly stated both in the front matter and within each chapter. This work does not, to the best of my knowledge, infringe on any copyright law or other Intellectual Property Rights, or contain confidential material. I accept the use of plagiarism detection software to check the electronic version of this thesis and confirm that it has not been previously submitted for the award of a degree by Imperial College London or any other university.

Details of Collaboration and Publications

Author contributions and publications are listed below. This work was supported by the Natural Environment Research Council (NERC) studentship NE/S007415/1, the Evolution and Education Trust (EET), the Pragnell Fund, and the Emily Holmes Scholarship. The PhD was supervised by Dr Ester Gaya (Royal Botanic Gardens Kew), Dr Tom Prescott (RBG Kew) and Professor Tim Barraclough (Oxford University and Imperial College London). For all sections that include work done in collaboration with others, I state the manner of collaboration in the relevant methods section and subsequently use the 'we' pronoun for methods and results.

Chapter 2 - This chapter was published in Genome Biology and Evolution (Llewellyn et al. 2023b). **Theo Llewellyn** performed the molecular lab work, designed and implemented the bioinformatics analysis and wrote the paper. Ester Gaya, Tom Prescott and Tim Barraclough supervised the work, designed the analysis and provided feedback on writing. Xavier Llimona, Marina Temina and Andre Aptroot collected lichen samples. Reuben W. Nowell assisted with designing the metagenome assembly pipeline and Pfam seed alignment analysis.

Chapter 3 - **T.L.** performed the molecular lab work, designed and implemented the bioinformatics analysis and wrote the chapter. E.G. and T.B. supervised the work, designed the analysis and provided feedback on writing. X.L., M.T., A.A., Alejandro Huereca Delgado, Jorge Guillermo-Guzman, Manuel L. Gil, E. Déniz, Erik Möller and Annie Evankow collected lichen samples. Sandra Alvarez-Carretero assisted with designing MCMCTree time-calibration analysis.

Chapter 4 - **T.L.** performed the molecular and chemical lab work, designed and implemented the bioinformatics analysis and wrote the chapter. E.G., T.P., and T.B. supervised the work, designed the analysis and provided feedback on writing. Pierre

Le Pogam Alluard performed mass spectrometry and provided accompanying methods. P.LPA. and Damien Olivier-Jimenez assisted with designing metabolome data processing, feature identification, and molecular networking pipeline.

Chapter 5 - **T.L.** performed the molecular and chemical lab work, fungal culturing, designed and implemented the bioinformatics analysis and wrote the chapter. T.P., T.B. and E.G. supervised the work, designed the analysis and provided feedback on writing. Rui Fang assisted with VelosPro LC-MS and XCalibur feature detection. Anthony Maxwell performed yeast Topo II assay and assisted with writing accompanying methods. Valerij Talagayev and Gerhard Wolber assisted with designing topoisomerase II structure modelling pipeline and V.T. implemented Alphafold on behalf of **T.L**.

Appendix - Appendix A.5 was published in Genome Biology and Evolution (Llewellyn et al. 2023a). **T.L.** performed the molecular lab work, designed and implemented the bioinformatics analysis and wrote the paper. **T.L.** and Sahr Mian performed flow cytometry lab work. Rowena Hill performed protein annotation. E.G., S.M., R.H., and Ilia Leitch provided feedback on writing.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Contents

Abstract 1

Acknowledgments 2

Statement of Originality 3

Details of collaboration and publications 4

Copyright Declaration 6

Contents 7

List of Figures 9

List of Tables 12

List of Abbreviations 14

1 Introduction **16**

1.1 The evolution of fungal secondary metabolism 16

1.2 Secondary metabolism in lichen-forming fungi 21

1.3 The *Teloschistales* as a case study for exploring secondary metabolites as adaptive traits 24

1.4 Thesis outline 28

2 Metagenomics Shines Light on the Evolution of 'Sunscreen' Pigment Metabolism in the *Teloschistales* (Lichen-Forming *Ascomycota*) **30**

2.1 Abstract 31

2.2 Introduction 31

2.3 Materials and Methods 35

2.4 Results 41

2.5 Discussion 51

2.6 Conclusions 59

2.7 Acknowledgements 59

2.8 Data availability statement 60

3 Phylogenomics and molecular systematics of the *Teloschistaceae* 61

3.1 Abstract . . . 61

3.2 Introduction . . . 62

3.3 Materials and Methods . . . 65

3.4 Results . . . 73

3.5 Discussion . . . 82

3.6 Conclusions . . . 95

4 Paired genome-metabolome analysis reveals a complex interplay between secondary metabolite profiles, biosynthetic gene cluster variation and phylogenetic relatedness 97

4.1 Abstract . . . 97

4.2 Introduction . . . 98

4.3 Materials and Methods . . . 100

4.4 Results . . . 109

4.5 Discussion . . . 119

5 Lichens have evolved self-resistance mechanisms to withstand the antifungal effects of anthraquinones 125

5.1 Abstract . . . 125

5.2 Introduction . . . 126

5.3 Methods . . . 129

5.4 Results . . . 136

5.5 Discussion . . . 144

6 Summary & Conclusions 149

References 155

A Appendices 192

A.1 Chapter 2 Supplementary Materials . . . 193

A.2 Chapter 3 Supplementary Materials . . . 199

A.3 Chapter 4 Supplementary Materials . . . 219

A.4 Chapter 5 Supplementary Materials . . . 225

A.5 *Ramalina farinacea* genome note (doi:10.1093/gbe/evad074) . . . 227

List of Figures

1.1 A schematic depicting the biosynthetic pathway of the polyketide parietin . 18
1.2 A visual overview of lichen secondary metabolite diversity. 23
1.3 Photographs showing lichen-dominated arid landscapes and examples of lichen growth forms and ecologies 25

2.1 Genome-scale maximum likelihood phylogeny including representatives of the largest class of lichenised fungi, the *Lecanoromycetes* (*Ascomycota*). . . 44
2.2 Diversity and distribution of secondary metabolite biosynthetic gene clusters (BGCs) in the *Lecanoromycetes*. 47
2.3 BiG-SCAPE similarity network, presence-absence matrix and PKS maximum-likelihood gene tree for putative anthraquinone BGC families. 49
2.4 Anthraquinone BGC structure and PT domain phylogeny 52

3.1 Number of described species vs genera in the top 50 most diverse lichen families 65
3.2 Time-calibrated maximum likelihood species tree of 550 *Teloschistales* and two *Caliciales* lichen-forming fungi. 75
3.3 Comparison of mean node ages and 95% HPD confidence intervals between MCMCTree and RelTime using 84-taxon genome-scale data. 76
3.4 Comparison of 95% HPD confidence interval size for node ages estimated with multi-locus data in Gaya et al. (2015) and genome-scale data in this study. 77
3.5 552-taxon multilocus phylogeny showing higher taxon delimitation by integrative taxonmy and GMYC 81

4.1 Comparison of untargeted metabolomics analysis of 79 lichen-forming fungi ordered by either hierarchical clustering of metabolite presence-absence or ML phylogeny. 111

4.2 PCoA plots of BGC family and LCMS computed neutral metabolites presence-absence patterns in the *Teloschistales*. 112
4.3 Fritz and Purvis' D-statistic results for BGC families and secondary metabolites for the 79-taxon dataset. 115
4.4 Anthraquinone subnetworks from GNPS feature-based molecular networking. 116
4.5 Synteny plot showing the organization of anthraquinone clusters in the *Teloschistales*. 117
4.6 Ancestral state reconstruction of genes within anthraquinone BGC. 118

5.1 Gel electrophoresis photos showing results of relaxation assay of lichen compounds with yeast DNA topoisomerase II system. 137
5.2 Presence and concentration of emodin and parietin in cultured aposymbiotic lichen mycobionts and whole lichen thalli 138
5.3 Fluorescence microscope images of lichen mycobiont cultures showing parietin autofluorescence. 139
5.4 Topoisomerase II protein domains and 3D structure showing sites under positive selection in the *Teloschistaceae* lineage 143

A.1 An example of mycobiont read filtering combining BlobTools and CONCOCT methods for the lichen metagenome of *Xanthoria mediterranea* (*Teloschistaceae*). 194
A.2 Violin plots comparing genome assembly statistics of published lichen mycobiont genomes from axenic culture with mycobiont genomes from metagenomic data. 196
A.3 Scatterplots showing the relationship between number of BGCs per genome and assembly fragmentation for 50 lichen mycobiont genomes. 197
A.4 Results of Interpro functional analysis showing gene annotations of the orthogroup containing putative anthraquinone PKSs. 198
A.5 Tanglegram comparing topology of ML phylogenies inferred from trimmed and untrimmed alignments. 213
A.6 Tanglegram comparing topology of the concatenated ML phylogenomic tree inferred by RAxML with the coalescent-based ASTRAL tree. 214
A.7 84-taxon ML genome-scale tree inferred using 1254 loci. 215
A.8 Time-calibrated phylogeny of 550 *Teloschistales* taxa and *Caliciales* outgroup showing MCMCTree 95% HPD confidence intervals. 216

A.9 Histogram showing the node ages of *Teloschistales* genera as estimated by MCMCTree. 217
A.10 Mean MCMCTree posterior vs mean prior node ages for the 84-taxon genome-scale phylogeny. 218
A.11 Presence-absence matrix of annotated secondary metabolites from GNPS feature-based molecular networking. 220
A.12 PCoA results of BGC and LCMS computed neutral metabolite presence-absence patterns in the *All Taxa* lichen dataset. 223
A.13 PCoA results of BGC presence-absence patterns in the *Teloschistales* lichen dataset with individual BGCs fitted as environmental variables. 223
A.14 Results of enrichment analysis testing whether *Teloschistales* taxa are more enriched in anthraquinone metabolites. 224
A.15 Concentrations of emodin and parietin in cultured aposymbiotic lichen mycobionts and whole lichen thalli. 226

List of Tables

2.1 Genome assembly statistics for 24 newly sequenced and assembled *Lecanoromycetes* mycobiont genomes. 42

2.2 Results of BiG-SCAPE clustering of BGCs into BGC families. 45

3.1 Node calibrations used for the time calibration of the 84-taxon genome-scale tree. 70

4.1 Mantel test statistics based on Pearson's product-moment correlation testing correlation between BGC family presence-absence distance matrix and metabolite distance matrices. 113

4.2 Results of Blomberg's K and Pagel lambda (λ) analysis. 114

5.1 Likelihood ratio tests (LRT) of branch and branch-site models. 140

5.2 CODEML branch-site model results showing proportion of amino acid sites in each site class and corresponding dN/dS (ω) values. 141

5.3 Bayes Empirical Bayes (BEB) analysis of sites under positive selection in the *Teloschistaceae* foreground lineage for the DNA repair genes encoding Hsp90 and topoisomerase II. 142

A.1 Collection information for 24 newly sequenced lichens in Chapter 2. 193

A.2 Lichen genome assemblies in the custom *Lecanoromycetes* database used in BLAST, phylogenomics and comparative genomics analysis. 195

A.3 Collection information for 59 newly sequenced lichens in Chapter 3. 200

A.4 Multilocus *Teloschistales* sequences included in 552-taxon molecular phylogenetic analysis. 201

A.5 Query genes used to pull six marker loci from assembled lichen genomes. . . 211

A.6 Genome assembly statistics for the 59 new assembled mycobiont genomes in Chapter 3. 212

A.7 Collection information for eight newly sequenced lichens in Chapter 4. . . . 219

A.8 Assembly statistics for eight additional non-*Teloschistales* mycobiont genomes in Chapter 4. 219
A.9 AIC results comparing the fit of three alternative evolutionary models for BGC and metabolite number per sample. 219
A.10 hOUwie results of joint modelling number of anthraquinones and anthraquinone BGC gene evolution. 220
A.11 Lichen samples used for LCMS emodin and parietin detection/quantification. 225
A.12 Collection information for four newly sequenced lichens in Chapter 5. 225
A.13 Genome assembly statistics for additional non-*Teloschistales* mycobiont genomes in Chapter 5. 225

List of Abbreviations

ABC ATP-binding cassette.

ACP acyl-carrier protein.

AIC Akaike information criterion.

AT acyltransferase.

BEB Bayes Empirical Bayes.

BGC biosynthetic gene cluster.

BGCF biosynthetic gene cluster family.

BM Brownian motion.

CDS coding sequence.

CTD C-terminal domain.

DUF domain of unknown function.

GMYC generalised mixed yule coalescent.

hESU higher evolutionary significant unit.

HPD highest posterior density.

HPLC high-performance liquid chromatography.

IC inhibitory concentration.

ITS internal transcribed spacer.

KS ketosynthase.

LC-MS liquid chromatography mass spectrometry.

LDI-MS laser desorption/ionization time-of-flight mass spectrometry.

LRT likelihood ration test.

LSD least-square dating.

MβL-TE metallo-β-lactamase type thioesterase.

MAG metagenome-assembled genome.

MCMC Markov chain Monte Carlo.

MFS major facilitator superfamily.

ML maximum likelihood.

NRP non-ribosomal peptide.

NRPS non-ribosomal peptide synthase.

nucLSU nuclear ribosomal large unit.

OLS ordinary least-squares.

ONT Oxford Nanopore Technology.

OU Ornstein-Uhlenbeck.

PCoA principal coordinate analysis.

PERMANOVA permutational analysis of variance.

PKS polyketide synthase.

RiPP ribosomally synthesised and post-translationally modified peptide.

RPB1 RNA polymerase II largest subunit.

RPB2 RNA polymerase II second largest subunit.

RRF relative rate framework.

RT retention time.

SM secondary metabolite.

SSU small subunit.

TC terpene cyclase.

TLC thin layer chromatography.

UV ultraviolet.

Chapter 1

Introduction

Publication details

Parts of this chapter have been published in the introduction of the following paper:

> **Llewellyn, T.**, Nowell, R., Aptroot, A., Temina, M., Prescott, T.A.K., Barraclough, T.G., Gaya, E. (2023). Metagenomics Shines Light on the Evolution of "Sunscreen" Pigment Metabolism in the *Teloschistales* (Lichen-Forming *Ascomycota*), *Genome Biology and Evolution*, 15(2):evad002. DOI: 10.1093/gbe/evad002

T.L. wrote the introduction. E.G., T.P. and T.B. supervised the work and provided feedback on the writing.

1.1 The evolution of fungal secondary metabolism

Organisms interact with their environment via their phenotypic traits. If these traits facilitate reproduction and survival, they are termed adaptive traits. In animals, adaptive traits are predominantly morphological, such as a protective exoskeleton to defend against predators, or behavioural, like burrowing to avoid harsh climatic conditions. In contrast, fungal adaptive traits are often biochemical. Fungi secrete a remarkable diversity of chemical compounds that mediate their interactions with other organisms and the abiotic environment. Toxins inhibit competitors and deter grazers (Rohlfs 2015; Boddy et al. 2017), virulence factors allow pathogens to evade host defences and manipulate host

behaviour (Williams et al. 2011), and pigments protect against solar radiation and desiccation (Bills et al. 2016). These low molecular weight compounds are typically classified as secondary or specialised metabolites as they are not essential for cell growth.

Fungal secondary metabolites display a wide range of biological activities and are some of the most potent bioactive compounds in nature. These bioactive properties have long been known by humans and have had both revolutionary and damaging effects on human society. On the one hand, fungal-derived drugs, such as the β-lactam antibiotic penicillin and the immunosuppressant cyclosporine, have benefitted global health immeasurably (Fleming 1929; Chain et al. 1940; Bills et al. 2016; Prescott et al. 2023). On the other hand, dangerous levels of fungal mycotoxins contaminate around 25% of global food crops (Eskola et al. 2020), resulting in severe poisoning and devastating economic losses (Strosnider et al. 2006; Jallow et al. 2021). Additionally, mycotoxins in indoor environments are a major cause of respiratory illness (Bräse et al. 2009).

Given their many applications, bioactive fungal secondary metabolites have been tested and catalogued in detail, with over 15,000 having been described thus far (Bills et al. 2016). However, historically, research has been biased towards 'useful' compounds, and only a fraction of the known fungi and their associated compounds have been assessed. Therefore, the true diversity of fungal secondary metabolites is undoubtedly much higher (Prescott et al. 2023).

Known fungal secondary metabolites are classified according to their molecular structure or starter units. The main classes are non-ribosomally and ribosomally synthesised peptides (NRPs and RiPPs), polyketides, terpenoids, and shikimate-derived compounds (Bills et al. 2016). Of these, polyketides are the most abundant and diverse. Most fungal polyketides are synthesised via the acetate polymalonyl pathway, a series of condensation reactions catalysed by a polyketide synthase (PKS) enzyme (Figure 1.1). Fungal PKSs are typically single multifunctional proteins with multiple catalytic domains (Type I PKSs) (Crawford et al. 2010).

Biosynthetic pathways for fungal secondary metabolites often require the concerted effort of multiple enzymes. The genes encoding these enzymes are not randomly distributed across the genome in fungi. Instead, genes that participate in the same biosynthetic pathway tend to group together on the chromosome, forming biosynthetic gene clusters (BGCs). A single BGC can consist of core, tailoring and accessory genes. Core genes encode en-

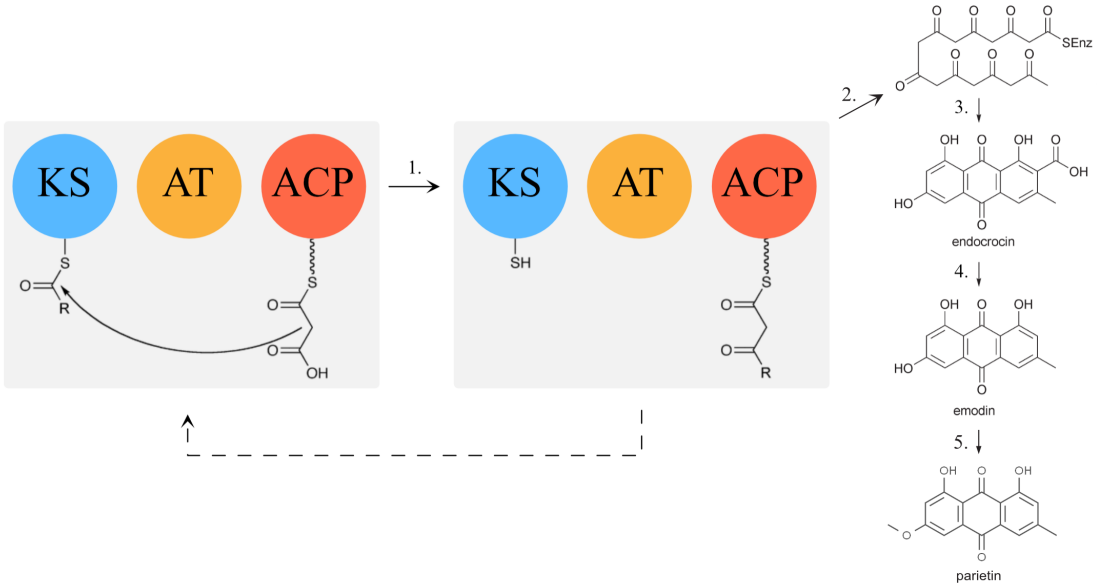


Figure 1.1: Biosynthetic pathway of the polyketide parietin. Grey blocks represent polyketide synthase (PKS) enzymes with three key domains: a ketosynthase (KS), an acyltransferase (AT) and an acyl-carrier protein (ACP). The starter ester unit acetyl-CoA is loaded into the ACP domain via the AT domain. 1) Starter units are extended by a condensation reaction with an extender unit at the KS domain, which adds a ketide unit to the chain. This is repeated cyclically until the length of the final polyketide has been reached. 2) Chain extension terminates, leaving a single polyketide chain attached to the enzyme (Enz). 3) The linear chain is cyclised to form an aromatic intermediate endocrocin. 4) The intermediate is decarboxylated to produce the emodin. 5) A methyl group is added to emodin by a methyltransferase enzyme to produce the final product, parietin. Biosynthetic pathway from Crawford et al. (2010) and Stocker-Wörgötter (2008).

zymes, such as polyketide synthases (PKS) or terpene cyclases (TC), that synthesise the metabolite skeleton. Products of tailoring genes, such as halogenases or methyltransferases, then modify the metabolite skeleton. Finally, accessory genes encode proteins for additional functions like metabolite transport and transcription regulation (Rokas et al. 2018). Although genes within a BGC are co-regulated, unlike bacterial operons, each one is individually transcribed.

Each BGC encodes one or more secondary metabolite products. However, BGCs producing the same metabolite in different species are not necessarily homologous, as there are often various possible biosynthetic pathways for the same metabolite. In fact, there are many known examples where BGCs producing the same metabolite have evolved independently in unrelated species, an example of convergent evolution (Slot et al. 2010; Elmore et al. 2015). The fact that these BGCs evolve repeatedly suggests that clustering genes into BGCs is an advantageous trait that can be driven by natural selection (Nei 1967). Rokas et al. (2018) discuss three genetic models in which natural selection could drive BGC formation and maintenance. Firstly, as clustered genes tend to be co-expressed

(Hurst et al. 2002), selection for tighter control of gene expression should promote BGC evolution. Secondly, selection for tighter clustering on the chromosome can reduce the likelihood of recombination disrupting interacting loci (Nei 1967; Wong et al. 2005). Thirdly, selection for clustering can be advantageous as it allows entire biosynthetic pathways to be horizontally transferred more easily (Wisecaver et al. 2015).

These models provide the genetic mechanisms through which selection promotes BGC evolution. However, selection does not act on the mechanism but on a phenotype. Whilst the phenotype for individual gene evolution is the function of a compound, the most likely phenotype for BGC evolution is toxicity avoidance (McGary et al. 2013). Fungal secondary metabolites are double-edged swords. Their potency comes from the fact that they commonly target universal cellular processes; however, this inadvertently targets the producing fungus too (Bills et al. 2016). Therefore, to prevent self-toxicity, the evolution of toxic compounds must be accompanied by the evolution of self-resistance mechanisms. Gene clustering provides a solution to this by combining self-resistance genes with toxin-producing genes. As mentioned previously, BGCs often encode transporter proteins. These act like efflux pumps, rapidly exporting metabolites as they are produced, which prevents toxic accumulation within the cell (Keller 2015). Tailoring genes like methyltransferases also promote toxicity avoidance by detoxifying harmful intermediates into less toxic derivatives (Scharf et al. 2010; Nielsen et al. 2013). Furthermore, BGCs can encode resistant duplicates of the enzyme that a toxin targets. These duplicates can contain point mutations that reduce toxin binding affinity or more simply soak up the toxin, further reducing toxic effects (Wiemann et al. 2013; Nyilasi et al. 2014).

Consequently, a clustered biosynthetic pathway including toxicity-avoidance genes is more resistant to self-poisoning than an unclustered one. McGary et al. (2013) provide empirical support for the toxicity-avoidance hypothesis, showing that biosynthetic pathways involving toxic compounds are more likely to be clustered than non-toxic ones. BGCs, therefore, offer a series of evolutionary advantages over non-clustered genes, which may explain their prevalence in fungal genomes. In contrast, BGCs in plants are much rarer (Wisecaver et al. 2017) and in animals, apart from some interesting exceptions (Moran et al. 2010; Scesa et al. 2022), are practically non-existent.

Fungal BGCs are not only numerous but highly variable. Studies have shown that BGCs are largely species-specific (reflecting variability in species chemistries) and impor-

tantly represent hotspots of evolution within the genome (Wisecaver et al. 2014; Koczyk et al. 2015; Theobald et al. 2018; Vesth et al. 2018; Gluck-Thaler et al. 2020). Lind et al. (2017) summarise five key genetic processes that can explain this variation:

1. Functional polymorphism where the same BGC evolves to produce different metabolites in different taxa

2. Positional polymorphism where BGCs vary in their position within the genome, resulting in changes to BGC regulation

3. Polymorphism in allelic content of a single gene within the BGC, which can alter protein function

4. Gains and losses of individual genes within the BGC resulting in changes in metabolite structure, expression or transport

5. Gains and losses of entire BGCs leading to gains or losses of metabolites

These processes are not mutually exclusive, and all five can be in effect within a single fungal genome. Together, they explain the high diversity in fungal BGCs that can be observed at phylum (Robey et al. 2021), class (Gluck-Thaler et al. 2020), genus (Vesth et al. 2018), and species level (Lind et al. 2017). This high rate of BGC evolution and variability between taxa also supports the role of secondary metabolites as effective adaptive traits, as it allows for adaptation to changing environmental conditions. However, perhaps more importantly, it allows for rapid adaptation in response to changing biotic interactions leading to coevolution.

Comparative genomics has revealed that BGCs do not occur evenly across the fungal tree of life. The early diverging *Chytridiomycota* and *Mucoromycota* phyla generally have much fewer BGCs than the *Ascomycota* and *Basidiomycota* (*Dikarya*) (Bills et al. 2016; Rokas et al. 2018; Robey et al. 2021). Within the *Ascomycota*, the *Pezizomycotina* subphylum is particularly rich, with some taxa having up to one hundred BGCs per genome. In contrast, unicellular yeasts of the subphyla *Saccharomycotina* and *Taphrinomycotina* have relatively impoverished BGC repertoires, suggesting BGC diversification in the *Ascomycota* occurred at the root of the *Pezizomycotina* (Rokas et al. 2018). Within the *Basidiomycota*, species in subphylum *Agaricomycotina* have the most diverse BGC repertoires (average of ∼14 BGCs per genome), albeit to a lesser extent than species in the

Pezizomycotina (average ∼40 BGCs per genome) (Robey et al. 2021). In addition, the distribution of BGC classes varies across the fungal phylogeny. For example, *Agaromycotina* have expanded Terpene and alkaloid BGC sets whilst *Pezizomycotina* have diversified more in PKS and NRPS BGCs (Rokas et al. 2018; Robey et al. 2021). This likely reflects ecological differences between these clades. For example, the abundance of alkaloids in mushroom-forming *Agaromycotina*, can be explained by the pressure they face from animal fungivores (Kosentka et al. 2013).

Comparing BGC repertoires across taxa is crucial for understanding how secondary metabolite innovation occurs and why different clades evolve different sets of BGCs. However, such comparisons are limited to the available fungal genomes, which are heavily biased towards certain fungal clades. A recent gap analysis by Hill et al. (2021) found that even for the *Ascomycota*, arguably the most well-studied fungal phylum, only around 50% of described orders had published genome assemblies. This lack of genomic data is especially apparent in clades with high proportions of mutualistic symbionts, such as endophytes and lichens.

1.2 Secondary metabolism in lichen-forming fungi

Lichens are obligate mutualisms between a primary fungus (mycobiont or lichen-forming fungus), one or more species of photosynthetic algae and/or cyanobacteria (photobiont), and various additional fungi and bacteria commonly referred to as the lichen microbiome (Lawrey et al. 2003; Nash 2008; Grube et al. 2009; Spribille et al. 2016; Hawksworth et al. 2020). Together they form stable perennial structures known as thalli. Much of the thallus is formed by mycobiont hyphae, which control water and gas exchange. Mycobionts can reproduce sexually and asexually via both specialised and non-specialised reproductive structures. Photobionts export photosynthates to the mycobiont and reproduce asexually. Cyanobacteria additionally provide nitrogen through atmospheric fixation. The role of the microbiome is still under debate, with some studies suggesting they actively contribute to the symbiosis (Grube et al. 2015; Spribille et al. 2016), whilst others posit a more commensal role akin to asymptomatic plant endophytes (U'Ren et al. 2010). Lichens also host a rich community of highly specialised lichen parasites termed lichenicolous fungi (Lawrey et al. 2003).

Lichen-forming fungi have arisen independently multiple times and encompass approximately 19,000 currently named species in 1000 genera (17% of known fungal diversity) (Lücking et al. 2017). Of these, almost 80% fall within class *Lecanoromycetes* O.E.Erikss. & Winka (*Pezizomycotina*, *Ascomycota*). Photobionts are much less diverse, belonging to around 100 described species across 50 genera (Sanders et al. 2021), although this is likely an underestimate (Muggia et al. 2020). This makes partner specificity highly imbalanced, with most mycobionts partnering with a small pool of photobiont species, whilst some predominant photobiont genera partner with thousands of mycobiont species (Sanders et al. 2021). This imbalanced pairing is partly due to the fact that mycobionts rely on the photobiont to survive, whereas the photobionts can survive in a free-living state. By specialising to partner with certain algal species mycobionts can obtain the maximum symbiotic benefit possible.

Lichens are well known for their secondary metabolite diversity, with over 1000 lichen compounds identified to date. Mycobiont-derived compounds make up almost all of these (Molnár et al. 2010) and are typically classified according to their biosynthetic pathways: polymalonate-derived, shikimate-derived, and mevalonate-derived compounds (Figure 1.2). As with non-lichenised *Ascomycota*, polymalonate-derived polyketides are the most abundant lichen secondary metabolites and include important compounds such as usnic acids, anthraquinones, depsides and depsidones (Stocker-Wörgötter 2008). Lichen metabolomes are also highly specialised, with only around 5% of lichen metabolites found in other organisms (Elix et al. 2008). This metabolic diversification and specialisation suggests secondary metabolites have been vital for adapting to new ecological niches and changing environments. These compounds are also assumed to be pivotal to survival given the significant proportion of biomass and energy that lichens dedicate to them, sometimes constituting up to 20% of a lichen's dry weight (Shrestha et al. 2013).

Lichen secondary metabolites are mainly deposited as extracellular crystals on the upper surface of the lichen thallus (especially the cortex). This is unusual for fungi and may reflect that lichens are constantly exposed to the aerial environment (an exceedingly rare habit in filamentous fungi) (Spribille et al. 2022). Functions of lichen secondary metabolites are manifold and include UV protection, anti-herbivory, antimicrobial activity, and metal homeostasis (Calcott et al. (2018) and references within).

More so than in other groups of macrofungi, secondary metabolites have shown to be

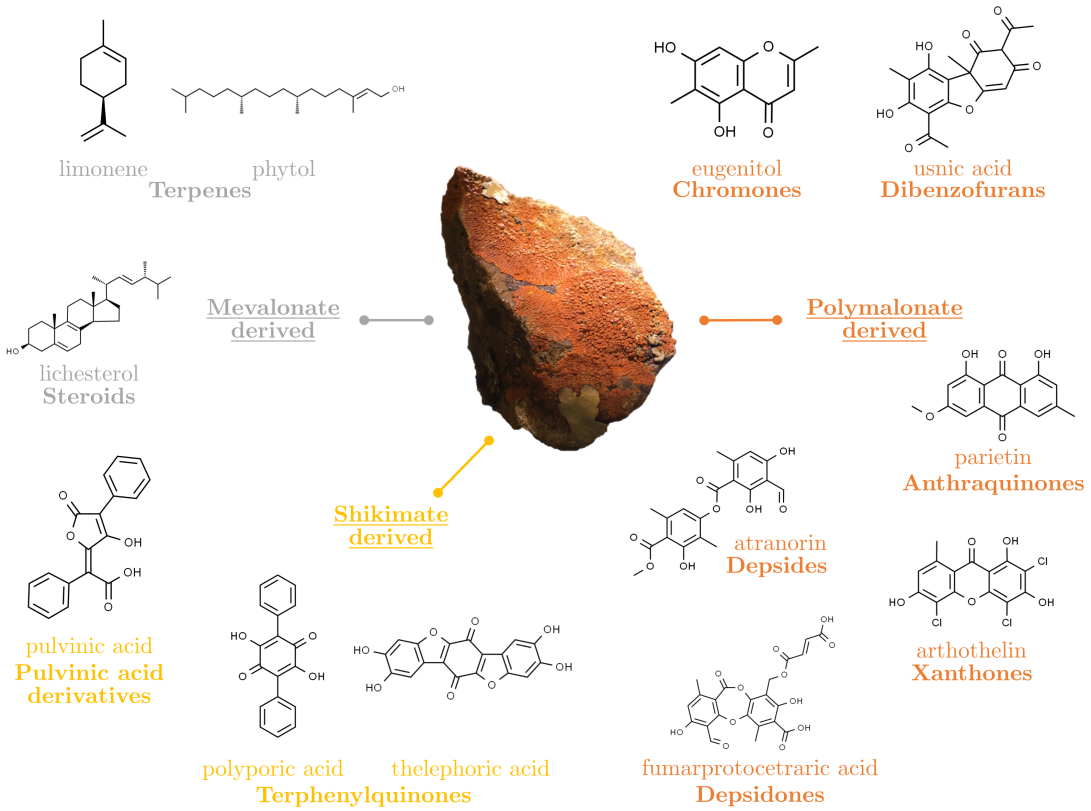


Figure 1.2: An overview of the chemical diversity of lichen secondary metabolites. Most lichen secondary metabolites are synthesised by one of three main biosynthetic pathways: the acetyl-polymalonate pathway (orange), the shikimate pathway (yellow), and the mevalonate pathway (grey). Compound classes are shown underneath compounds in bold text.

useful in lichen taxonomy and are routinely analysed for species identification and description. This has led to chemotaxonomy becoming a strong and central discipline of lichenology, with spot tests, TLC and HPLC regularly complementing morphological and molecular data. Taxonomic and systematic studies commonly report presence/absence of key metabolites (chemotypes) or more detailed chemical profiles known as chemosyndromes: constellation-like chemical profiles with one or two consistently produced compounds accompanied by structurally related, more ephemeral compounds. Chemosyndromes, a term first coined by lichenologists Culberson et al. (1976), significantly advanced lichen chemotaxonomy as they accounted for intraspecific variation, thus providing more accurate metabolite descriptions. However, given its high variation, the effectiveness of chemical data in taxon delimitation has proved irregular. Depending on the lichen group, integrating chemistry with phylogenetics shows contrasting results, with some groups showing strong correlations between metabolic similarity and phylogenetic relatedness (Miadlikowska et al. 2000; Tehler et al. 2001; Schmitt et al. 2004; Lücking et al. 2008; Crespo et al. 2010),

whilst others show little to no correlation (LaGreca 1999; Blanco et al. 2006; Nelsen et al. 2008; Lutsak et al. 2017).

Lichens have successfully colonised all continents and are well-known for their ability to survive in inhospitable or extreme ecosystems like deserts, tundra, and high-altitude mountain ranges (Lange 1990; Kappen 2000; Kranner et al. 2008). They are often the dominant biota in these ecosystems (Figure 1.3) and provide essential ecosystem services, including soil formation, nutrient and water cycling, and habitat provision. Their success in exposed habitats comes from various ecological, morphological and physiological adaptations. Firstly, lichens are desiccation-tolerant, allowing them to survive extreme temperature fluctuations and extended periods of drought (Kranner et al. 2008). Secondly, they can absorb moisture directly from the atmosphere, allowing them to grow in habitats with little to no rainfall. Additionally, absorbing water and nutrients from the atmosphere allows lichens to grow on nutrient-poor substrates like exposed rock, sand (Figure 1.3) and even man-made materials. Finally, their extracellular layer of secondary metabolites reflects or absorbs light, protecting the symbionts from solar radiation and UV-induced DNA damage (Solhaug et al. 2010). By reducing water loss, light reflection further protects against desiccation and extreme temperatures.

1.3 The *Teloschistales* as a case study for exploring secondary metabolites as adaptive traits

One lichen group that has been particularly successful in exposed ecosystems is the family *Teloschistaceae* Zahlbr. (*Lecanoromycetes*, *Ascomycota*). With over 1,000 species, the *Teloschistaceae* is one of the most speciose lichen families and though cosmopolitan in its distribution, has a particular association with exposed, arid regions (Arup et al. 2013; Lücking et al. 2017). The family is phenotypically and ecologically diverse but with a prevalence of rock-inhabiting crustose species. In contrast, its sister families *Megalosporaceae*, *Letrouitiaceae*, and *Brigantiaceae* are small tropical lineages dominated by epiphytes. During the Late Cretaceous, ancestral habitat switches from shaded epiphytic habitats to exposed rocky ones likely allowed the *Teloschistaceae* to radiate into some of the most arid ecosystems across the globe, which is where we currently find their biodiversity hotspots (Gaya et al. 2015). This habitat switch coincided with secondary



Figure 1.3: Photographs showing lichen-dominated arid landscapes and examples of lichen growth forms and ecologies. (*Top*) A lichen field dominated by *Teloschistes capensis* (L. f.) Müll. Arg. (*Teloschistaceae*, *Ascomycota*) in a coastal fog desert near Hentiesbaai, Erongo, Namibia. (*Bottom left*) *Dufourea flammea* (L. f.) Ach., a fruticose lichen (*Teloschistaceae*) growing directly on sand dunes in Namib-Naukluft National Park, Namibia. (*Bottom right*) *Stellarangia elegantissima* (Nyl.) Frödén, Arup & Søchting (*Teloschistaceae*), a crustose lichen growing on exposed quartz rock near Hentiesbaai, Erongo, Namibia. Photos by Theo Llewellyn June 2022.

metabolite innovation in the form of sun-screening anthraquinone production. This led Gaya et al. (2015) to hypothesise that anthraquinone evolution in the common ancestor of the *Teloschistaceae* may be linked to their evolutionary success in these ecosystems.

Anthraquinones are bright yellow, orange, and red aromatic polyketides that efficiently absorb UV and blue light (Solhaug et al. 1996; Solhaug et al. 2004). Their core struc-

ture consists of three fused benzene rings (anthracene) with two carbonyl groups (carbon double-bonded to oxygen) at positions 9 and 10 (see Figure 1.1 for a summary of the biosynthetic pathway). Whilst also found in some plants, animals, and non-lichenised fungi, we rarely observe such ubiquity and abundance of anthraquinones as in the *Teloschistaceae*. The first comprehensive analysis of *Teloschistaceae* anthraquinones was carried out by Santesson (1970) who identified 12 chemical groups based on the presence, absence and proportions of nine anthraquinones within *Caloplaca* Th.Fr. s.l., a historical genus that until recently contained almost half of the species in the family. Santesson's groupings were advanced by Søchting (1997), Søchting (2001) and Søchting et al. (2002) who identified nine additional anthraquinones and described 22 chemosyndromes for *Caloplaca*, *Teloschistes* Norman and *Xanthoria* (Fr.) Th.Fr. Currently, *Teloschistaceae* chemosyndromes are mainly informative at the species level and even for some genera. Subfamily-level patterns are more inconsistent, with chemosyndromes being relatively homogenous in the *Xanthoriodeae* but highly variable in the *Caloplacoideae* (Arup et al. 2013). Most *Teloschistaceae* species produce anthraquinones; however, a small number have secondarily lost them, instead producing structurally unknown grey or green pigments. Furthermore, species without anthraquinones do not form a single clade suggesting anthraquinones have been lost more than once (Frolov et al. 2020).

The principle function of anthraquinones in the *Teloscistaceae* is undoubtedly as photoprotective pigments (Solhaug et al. 1996; Solhaug et al. 2004). However, anthraquinones are also highly toxic, displaying potent antibacterial and antifungal properties (Agarwal et al. 2000; Hildebrandt et al. 2018). As a result, they are used as fungicides and pesticides (Hildebrandt et al. 2018). In animal systems, they also show antiproliferative and antimetastatic properties and are thus being explored as potential anticancer agents (Wu et al. 2007; Liu et al. 2011; Malik et al. 2021). Given their bioactive properties, *Teloschistaceae* lichens may also use anthraquinones as antimicrobial defences, or equally, they evolved as such and were then co-opted to act as photoprotective pigments. Furthermore, as *Teloschistaceae* fungi are constantly exposed to high concentrations of their own toxic anthraquinones, given the toxicity-avoidance hypothesis discussed previously, these fungi will likely have evolved self-resistance mechanisms to avoid suffering the toxic effects themselves.

We do not know how anthraquinones transitioned from a relatively rare, minor metabo-

lite trait in most lichens into a near-ubiquitous, dominant trait in the *Teloschistaceae*. This is due to two main factors: a lack of genomic data for the *Teloschistaceae* and its sister families (order *Teloschistales* D.Hawksw. & O.E.Erikss.) and the need for a robust phylogenetic tree with which to test evolutionary hypotheses.

Lichens have historically been under-represented in genomic studies. Their extremely slow growth makes it difficult to harvest enough material for genomic sequencing, and the complexities of isolating them from symbiosis make axenic culturing a challenge. In recent years, the application of metagenomic approaches has dramatically increased the number of available lichen-forming fungal genomes and metagenome-assembled genomes (MAGs) of lichen mycobionts now equal or even exceed the quality of those obtained from pure cultured material (Meiser et al. 2017; Greshake Tzovaras et al. 2020; McKenzie et al. 2020). Furthermore, thanks to long-read sequencing, we are beginning to see the first reference-quality lichen mycobiont MAGs (McKenzie et al. 2020; Wilken et al. 2020; Allen et al. 2021; Singh et al. 2021b; Gerasimova et al. 2022). Nevertheless, lichenised fungal genomes are still underrepresented in public repositories, with only 107 whole genome assemblies available for the *Lecanoromycetes*, compared to 1,711 and 3,600 for the *Dothideomycetes* and *Sordariomycetes* respectively, despite all three classes having similar numbers of described species (calculated using pipeline and dataset of (Hill et al. 2021)). Therefore, we must first address the lichen genomic data deficit to investigate hypotheses regarding anthraquinone innovation and secondary metabolism evolution in general.

A robust, comprehensively sampled phylogeny of the *Teloschistales* is the second missing factor for testing hypotheses about anthraquinone evolution. As with other fungi, the most common datasets for phylogenetic reconstruction of the *Teloschistales* consist of one or a few marker loci (Gaya et al. 2008; Gaya et al. 2012; Arup et al. 2013; Wilk et al. 2021). Although multilocus datasets allow for dense taxon sampling at a relatively low cost, they often struggle to resolve deeper divergences between genera and families. This lack of resolution has led to major uncertainty and disagreement regarding higher-level relationships in the clade. Genome-scale datasets can help to address this and are just beginning to be applied to lichens systematics (e.g. Leavitt et al. (2016), Pizarro et al. (2018), Grewe et al. (2020), and Díaz-Escandón et al. (2022)). Resolving deeper phylogenetic nodes is also invaluable for delimiting stable higher taxa, especially in clades like the *Teloschistales*, which lack reliable morphological characters.

1.4 Thesis outline

Lichens produce sophisticated arsenals of secondary metabolites, which allow them to thrive in extreme ecosystems across the globe. However, the genetic basis and evolutionary processes underpinning most lichen secondary metabolites have not been investigated. In this thesis, I focus on anthraquinone evolution in the *Teloschistales* as a case study to explore the questions: how do adaptive metabolic traits arise in lichen-forming fungi, and what are the evolutionary processes that shape metabolite diversification? I aim to answer these by addressing four main objectives:

Objective 1 - Characterise the biosynthetic genes responsible for anthraquinone production in the *Teloschistales*

Objective 2 - Produce a robust genome-scale time tree for the *Teloschistales* to resolve relationships between higher taxa and provide baseline data for phylogenetic comparative analysis

Objective 3 - Understand the evolutionary processes that shape metabolite diversification and determine how genomic changes affect the chemical phenotype

Objective 4 - Investigate self-resistance adaptations in the *Teloschistaceae* as a consequence of toxic anthraquinone production

Characterising biosynthetic gene clusters requires high-quality genome assemblies. In Chapter 2, I develop and implement a novel metagenomics pipeline to sequence, assemble and annotate *Teloschistales* genomes. Through comparative analysis of BGC repertoires, I test potential scenarios of BGC evolution in lichen-forming fungi. I also identify putative anthraquinone BGCs in *Teloschistales* genomes and compare their structure and gene content to BGCs in other fungi. With this, I investigate if *Teloschistales* anthraquinone BGCs contain unique elements that could explain the abundance and ubiquity of anthraquinones in the clade.

As mentioned above, a comprehensive phylogeny is a prerequisite for robust comparative trait evolution analysis. Therefore, in Chapter 3, I expand the metagenomics approach developed in Chapter 2 to sequence all major lineages in the *Teloschistales*. I use this dataset for a genome-level phylogenetic review of *Teloschistaceae* systematics, with

a particular focus on higher taxa delimitation. Given the breadth of the dataset, I also revisit divergence time estimation in the clade to understand the timing of *Teloschistales* evolution.

The relationship between genotype and metabolite phenotype can be complex. In Chapter 4, I generate novel untargeted metabolome data and use it alongside the genomic dataset and time-calibrated phylogeny generated in Chapter 2 and 3 to explore genomic and metabolic evolution across the whole clade. I use trait evolution methods to test whether genomic variation correlates with metabolic variation. I then focus specifically on anthraquinone BGCs and metabolites to explore potential scenarios of adaptive trait evolution in the *Teloschistales*.

Finally, as mentioned in Section 1.3, given the toxicity-avoidance hypothesis, I expect *Teloschistaceae* fungi to have evolved self-resistance mechanisms to anthraquinones. In Chapter 5, I first determine the cellular target of anthraquinones to understand how these compounds are toxic. I then investigate the role of tailoring enzymes in compound detoxification under both symbiotic and aposymbiotic conditions. I also test for positive selection in the genes encoding the cellular target of anthraquinones and search for unique mutations in these genes in the *Teloschistaceae*. Lastly, I use *in silico* protein modelling to understand whether these mutations may confer resistance to anthraquinone toxicity.

Chapter 2

Metagenomics Shines Light on the Evolution of 'Sunscreen' Pigment Metabolism in the *Teloschistales* (Lichen-Forming *Ascomycota*)

This chapter has been published as the following paper:

> **Llewellyn, T.**, Nowell, R., Aptroot, A., Temina, M., Prescott, T.A.K., Barraclough, T.G., Gaya, E. (2023). Metagenomics Shines Light on the Evolution of "Sunscreen" Pigment Metabolism in the *Teloschistales* (Lichen-Forming *Ascomycota*), *Genome Biology and Evolution*, 15(2):evad002. DOI: 10.1093/gbe/evad002

T.L. performed the molecular lab work, designed and implemented the bioinformatics analysis and wrote the paper. E.G., T.P. and T.B. supervised the work, designed the analysis and provided feedback on the writing. M.T. and A.A. collected lichen samples. R.N. assisted with designing the metagenome assembly pipeline and Pfam seed alignment analysis.

2.1 Abstract

Fungi produce a vast number of secondary metabolites that shape their interactions with other organisms and the environment. Characterising the genes underpinning metabolite synthesis is therefore key to understanding fungal evolution and adaptation. Lichenised fungi represent almost one-third of *Ascomycota* diversity and boast impressive secondary metabolites repertoires. However, most lichen biosynthetic genes have not been linked to their metabolite products. Here we used metagenomic sequencing to survey gene families associated with the production of anthraquinones, UV-protectant secondary metabolites present in various fungi, but especially abundant in a diverse order of lichens, the *Teloschistales* (class *Lecanoromycetes*, phylum *Ascomycota*). We successfully assembled 24 new, high-quality lichenised fungal genomes de novo and combined them with publicly available *Lecanoromycetes* genomes from taxa with diverse secondary chemistry to produce a whole-genome tree. Secondary metabolite biosynthetic gene cluster (BGC) analysis showed that whilst lichen BGCs are numerous and highly dissimilar, core enzyme genes are generally conserved across taxa. This suggests metabolite diversification occurs via re-shuffling existing enzyme genes with novel accessory genes rather than BGC gains/losses or de novo gene evolution. We identified putative anthraquinone BGCs in our lichen dataset that appear homologous to anthraquinone clusters from non-lichenised fungi, suggesting these genes were present in the common ancestor of the subphylum *Pezizomycotina*. Finally, we identified unique transporter genes in *Teloschistales* anthraquinone BGCs that may explain why these metabolites are so abundant and ubiquitous in these lichens. Our results support the importance of metagenomics for understanding the secondary metabolism of non-model fungi such as lichens.

2.2 Introduction

Fungi produce a high diversity of secondary metabolites, with functions including toxins (Boddy et al. 2017; Rohlfs 2015), virulence factors (Williams et al. 2011; Boyce et al. 2019), and protection against stressful environments (Bills et al. 2016). Many of these compounds are used by humans as powerful bioactive molecules. Understanding the genetic basis for metabolite production and how evolution acts on those genes to generate chemical diversity

is an important step for understanding fungal diversity and adaptation and for applications of biomolecules.

As discussed in Chapter 1.1, genes involved in the same secondary metabolic pathway in fungi often cluster into biosynthetic gene clusters (BGCs). BGCs offer a series of evolutionary advantages over non-clustered genes, including tighter control of expression (Hurst et al. 2002), reduced likelihood of recombination disrupting interacting loci (Nei 1967; Wong et al. 2005), and prevention of accumulation of self-toxic products through metabolite transporters and tailoring genes within the cluster (Scharf et al. 2010; Nielsen et al. 2013; McGary et al. 2013). Investigating variation in gene cluster content and structure across different taxa may therefore help us to understand the diversity and distribution of fungal secondary metabolites and the evolutionary processes that led to their formation (Slot et al. 2019).

The shift from sequencing individual genes to whole genomes has allowed us to study the full BGC repertoire of a fungus rather than just a few of its core biosynthetic genes. Results have shown that BGCs are diverse regions of the genome, with high variability even between closely related taxa (Wisecaver et al. 2014; Koczyk et al. 2015; Theobald et al. 2018; Vesth et al. 2018; Gluck-Thaler et al. 2020). This variability can be explained by multiple genetic processes, including polymorphism in BGC functionality, polymorphism in position in the genome, polymorphism in allelic content, gains/losses of individual genes, and gains/losses of whole BGCs (Lind et al. 2017). Comparing the increasing number of fungal whole genome sequences provides a way to test the contribution of these processes within a phylogenetic context. Major groups of fungi remain, however, relatively under-explored.

With approximately 19,000 currently named species, lichen-forming fungal symbionts represent almost one-third of the known *Ascomycota* (Lücking et al. 2017) and boast an impressive repertoire of secondary metabolite compounds (>1,000 compounds described to date (Huneck et al. 1996; Stocker-Wörgötter 2008; Elix 2014)), including molecules that allow them to inhabit harsh environments like exposed rock surfaces. However, due to practical challenges with axenic culturing and sequencing lichens, they are often under-represented in studies on BGC diversity and evolution. In recent years, metagenomics has helped to address this lack of genomic data and high-quality metagenome-assembled genomes (MAGs) of lichen mycobionts are now increasingly common (Meiser et al. 2017;

Greshake Tzovaras et al. 2020; McKenzie et al. 2020; Wilken et al. 2020; Allen et al. 2021; Singh et al. 2021b; Gerasimova et al. 2022). Nevertheless, as discussed in Chapter 1.3, lichen-forming fungal clades are still inadequately represented in public repositories compared to similar-sized non-lichen clades (Hill et al. 2021). For the lichen-forming fungi that have been sequenced, comparative analysis has shown that their BGCs are as numerous as in other classes of the *Ascomycota*, if not more so (Armaleo et al. 2011; Bertrand et al. 2018c; Calchera et al. 2019; Pizarro et al. 2020; Gerasimova et al. 2022). Additionally, these studies consistently show that for a given lichen species, the number of BGCs in the genome greatly outnumbers the known metabolites, suggesting that many secondary metabolites remain undiscovered.

A key challenge for genomic prospecting is linking genes or gene clusters to their metabolite products. In lichens, direct confirmation of gene function via genetic manipulation is hard because of their extremely slow growth and the complexity of maintaining them under aposymbiotic conditions. Instead, most studies adopt a homology-based approach where lichen genes of unknown function are mapped to genes that have been experimentally characterised in model fungi (Bertrand et al. 2018a). This approach has been used to suggest putative genes for the lichen compounds atranorin (Gerasimova et al. 2022), olivetoric and physodic acids (Singh et al. 2021b), and gyrophoric acid (Singh et al. 2022), amongst others. One-to-one mapping is not always possible because of extensive duplication in lichen metabolite genes (Opanowicz et al. 2006; Muggia et al. 2008; Schmitt et al. 2005), production of multiple compounds of the same class, or production of different metabolites by regulation of a single BGC (Singh et al. 2021b; Singh et al. 2022). Nonetheless, genome analyses provide a starting point for predicting metabolite repertoires, which in two cases have been confirmed by subsequent heterologous expression in non-lichen hosts (Kealey et al. 2021; Kim et al. 2021).

In this chapter, I use a comparative metagenomic survey approach to search for gene clusters associated with the production of a key lichen metabolite group anthraquinones in the diverse *Teloschistales* order. Anthraquinones are photoprotective pigments deposited in the upper cortex of lichens that efficiently absorb UV and blue light (Solhaug et al. 1996; Gauslaa et al. 2003). They are products of iterative multidomain PKS enzymes (Type I PKS/PKSI). Type I PKSs are often characterised by their thioesterase (TE) domain structure, the domain responsible for cleaving the polyketide from the PKS (Li et al.

2010b). In the case of fungal anthraquinones, their corresponding gene clusters consistently show a PKS that lacks a TE-domain adjacent to a distinct metallo-β-lactamase type TE gene (Awakawa et al. 2009; Chiang et al. 2010; Lim et al. 2012; Neubauer et al. 2016; Throckmorton et al. 2016; Griffiths et al. 2016). BGCs for different anthraquinones in non-lichenised fungi show a high level of sequence similarity and synteny suggesting anthraquinone production in these fungi is homologous. Whether this extends to *Teloschistales* fungi that produce much higher levels of anthraquinones has yet to be tested. Only two clusters have been putatively linked to anthraquinone biosynthesis in lichens (Wang et al. 2014; Bertrand et al. 2018a). While these show apparent homology to non-lichenised fungi, comparisons of more species are needed to confirm homology and explore anthraquinone BGC diversity across all lichen groups. A comparative approach will allow us to investigate how anthraquinone production transitioned from the relatively rare, minor lichen metabolite trait seen in most lineages into the dominant secondary metabolite and a near-ubiquitous trait in the *Teloschistales*.

We still know very little about the BGCs underpinning anthraquinone biosynthesis in lichens, particularly in the *Teloschistales*, for which only three full genome sequences are publicly available. In this investigation, I increased the number of available lichen genomes by implementing a metagenomics approach to sequence and assemble 24 new lichen genomes, with a special emphasis on the *Teloschistales*. These genomes were then used alongside 21 published *Lecanoromycetes* genomes to perform a comparative genomic analysis of lichen BGCs to predict links between these clusters and key chemical compounds. I asked whether lichen anthraquinone BGCs are homologous to those of non-lichenised fungi, whether *Teloschistales* BGCs contain unique elements that could explain anthraquinone prevalence in this clade, and whether metabolite diversification in lichens occurs through gains and losses of entire BGCs or shuffling of components within BGCs. Using anthraquinones as a case study, I explore how fungal secondary metabolite innovation occurs and hypothesise a genetic mechanism for how lichenised clades such as the *Teloschistales* specialise to produce such high concentrations of compounds.

2.3 Materials and Methods

DNA extraction & whole genome sequencing

For newly sequenced lichen genomes, sampling was focused on the order *Teloschistales*. I selected 22 samples covering 20 species to cover the three subfamilies included within the largest family *Teloschistaceae* and its sister family *Letrouitiaceae*. In addition, two non-*Teloschistales* taxa (*Diploschistes diacapsis* and *Umbilicaria vellea*) were sampled. Supplementary Table A.1 includes locality and collection information for all samples. One sample of note is *Teloschistes peruensis* (Ach.) J.W. Thomson (family *Teloschistaceae*). The species was located along a small ridge crest (∼100 x 30 m) of sedimentary uplift superimposed with powdery desert loess and cryptobiotic crust in the Lomas Amara in Peru. Another such habitat has yet to be found and the climatic, geological and topographic niches are clearly very unusual under today's conditions, possibly due to contraction of such climatic niche habitats and ecological degradation of other fog-oases (O. Whaley pers. comm.). The core area of Lomas de Amara is incorporated and protected under national designation as an 'Ecosistema Fragile' (Resolution No. 153-2018-MINAGRI-SERFOR) and *Teloschistes peruensis* has been assessed as critically endangered on the IUCN Red List (Ramos et al. 2021).

Lichen thalli that were clear from visible signs of other fungal contaminants were cleaned of substrate debris, washed with double-distilled water and transferred to 2 ml Eppendorf tubes. Samples were ground to a fine powder using a Mixer Mill MM 400 (Retsch, Germany) and two sterile stainless-steel beads per tube. DNA was extracted using a modified CTAB protocol (Cubero et al. 2002) or alternatively the Qiagen DNeasy® Plant Mini Kit following the manufacturer's protocol (Qiagen, Redwood City, CA, United States). Extractions were quality checked using a Quantas™ Fluorometer (Promega, UK) and NanoDrop™ 2000 Spectrophotometer (ThermoFisher Scientific, UK). Genome libraries were prepared using either the TruSeq Nano PCR-free DNA Kit (Illumina) or Nextera XT DNA Library Preparation Kit with a 550bp insert size. Paired-end libraries (2 x 151bp) were sequenced using a NovaSeq 6000 platform (Illumina, San Diego CA). Library prep and sequencing were performed at Macrogen (Macrogen Inc. South Korea).

Metagenome assembly and mycobiont genome isolation

Illumina paired-end reads were quality-checked using FASTQC v0.11.2 (Andrews et al. 2015). Low-quality bases and adapters were removed using Trimmomatic v0.36 (Bolger et al. 2014) with a four-base sliding window and minimum Phred score of 15. Only reads with a mate pair post-trimming were used for assembly. Reads were assembled de novo into contigs using two assembly tools for comparison, MEGAHIT v1.2.9 (Li et al. 2015) and MetaSPAdes v3.15.1 (Nurk et al. 2017), both with default settings. Total assembly size, number of contigs >500 base pairs, and N50 were calculated using QUAST v5.0.2 (Gurevich et al. 2013). Genome completeness as measured by gene sets was quantified using BUSCO v4.0.2 (Manni et al. 2021) and the BUSCO Ascomycota lineage gene set – ascomycota_odb10.2019-11-20 (1,706 single-copy orthologs, `https://busco.ezlab.org/list_of_lineages.html`). QUAST and BUSCO metrics were used to choose the best of the two assemblies (i.e., MEGAHIT vs. MetaSPAdes) for subsequent steps.

Reads belonging to the *Lecanoromycetes* mycobiont were isolated from the metagenome using a new pipeline presented here consisting of a combination of the BlobTools v1.1.1 workflow (Laetsch et al. 2017) and the metagenomic binning tool CONCOCT v1.1.0 (Alneberg et al. 2014). First, for BlobTools, read coverage was calculated using the function 'bbmap' in BBTools (Bushnell 2014) and then converted to a BlobTools 'covfile' using a one-line custom perl script (`https://github.com/theo-llewellyn/TeloschistalesMetagenomics/blob/main/mycobiont_filtering/BlobTools/bbmap.sh`). Metagenome contigs were then taxonomically identified using a double blast approach: 1) a DIAMOND BLASTx (Buchfink et al. 2014) against the Uniref90 database (Bateman 2019), and 2) a BLAST+ BLASTn (Camacho et al. 2009) against a custom database of all published genomes classified under *Lecanoromycetes* available in NCBI and JGI Mycocosm. Both blasts used an e-value cut-off of $1 * 10^{-25}$. BlobTools was then used to visualise the read coverage, GC-content, and taxonomic identification (phylum level) of all contigs. For the CONCOCT metagenome binning step, all metagenome-assembled contigs were first cut into 10K bp sections, and then initial raw reads were mapped against these 10K bp sections before finally binning the sections into metagenome-assembled genomes (MAGs) using default settings. Results of CONCOCT binning were visualised by plotting the read-coverage, GC-content, and

MAG identity for each contig using ggplot in R v.4.0.3 (Wickham 2016; R Development Core Team 2019).

The CONCOCT results and BlobTools identifications for each metagenome were overlapped to detect which MAGs had been identified as *Ascomycota* by BlobTools. In most cases, the mycobiont genome formed a linear cloud of contigs with 30% to 50% GC-content at a consistent coverage, a pattern also observed in other lichen metagenome studies (McKenzie et al. 2020; Tagirdzhanova et al. 2021). CONCOCT tended to split the linear cloud into one large metagenome bin at 50% GC-content and a series of smaller bins along the 'tail' of the linear cloud. These 'tail' bins were therefore merged with the large bin to form a single mycobiont MAG, following the example of Tagirdzhanova et al. (2021).

To remove contaminants, the putative mycobiont MAG was then subjected to a second round of the BlobTools workflow, where any contigs that did not have a top blast hit of either *Ascomycota* or a 'no hit' at all were removed. This was done using the BlobTools blobDB.bestsum.table.txt file as input for SeqKit's grep function (Shen et al. 2016). Supplementary Figure A.1 gives a visual representation of the mycobiont filtering pipeline. With contaminants removed, reads mapping to the remaining mycobiont contigs were extracted using the 'outm=filtered_R#.fq.gz' function of BBTools 'bbmap' (Bushnell 2014). QUAST and BUSCO were used to assess the quality of the final MAGs as described above.

Finally, redundans v0.14a (Pryszcz et al. 2016) was used to remove redundant contigs, scaffold the remaining contigs and close gaps. This produced my final *Lecanoromycetes* mycobiont MAGs, henceforth referred to as mycobiont genomes. These newly produced mycobiont genomes were then combined with 21 already published *Lecanoromycetes* mycobiont genomes downloaded from NCBI and Mycocosm (see Supplementary Table A.2 for details). I thank Professor Paul Dyer and Dr Igor Grigoriev for granting me permission to use the unpublished JGI Mycocosm genomes of *Xanthoria parietina* and *Usnea florida*. All subsequent analyses were performed on the 45-genome combined dataset of my 24 newly sequenced genomes and the 21 downloaded genomes.

Mycobiont gene prediction and orthologue identification

Repeat content libraries were generated de novo for each genome using RepeatModeler v2.0.1, implementing default settings with LTR discovery (Flynn et al. 2020) and then used

as a custom library for softmasking with RepeatMasker v4.1.0 (Smit et al. 2015). Gene prediction was performed on the masked assemblies using the funannotate v1.7.0 pipeline (Palmer et al. 2019). Funannotate performs gene prediction using Augustus (Stanke et al. 2006), GlimmerHMM (Majoros et al. 2004), and SNAP (Korf 2004) and consolidates results into a single set using EvidenceModeler (Haas et al. 2008). Expressed sequence tags (ESTs) and protein models from the genomes of *Xanthoria parietina*, *Cladonia grayi*, and *Usnea florida* were downloaded from JGI Mycocosm (Grigoriev et al. 2014; Armaleo et al. 2019)(Supplementary Table A.2) and provided to EvidenceModeler to inform gene prediction further. Repeat masking and gene prediction were not performed on the Mycocosm data as this information was already available. Orthologous genes (orthogroups) were identified from amino acid sequences of all 45 genomes using Orthofinder v2.4.0 with default settings (Emms et al. 2019). Single-copy orthogroups present in at least 75% of the taxa were used for subsequent tree reconstruction producing a final dataset of 2,214 orthogroups.

Lecanoromycetes phylogeny

The amino acid sequences for each of the 2,214 orthogroups were aligned using MAFFT v7.271 with the '-auto' option (Katoh et al. 2009). Ambiguously aligned regions were trimmed using TrimAl v1.4.rev12 with the '-automated1' option (Capella-Gutiérrez et al. 2009). Concatenation and coalescent-based approaches were compared to produce the *Lecanoromycetes* species tree. IQTree v2.0.6 (Nguyen et al. 2015) was used to produce a concatenated alignment of the trimmed multiple sequence alignments. The optimal amino acid substitution model for each orthogroup was identified using IQTree's inbuilt tool ModelFinder (Kalyaanamoorthy et al. 2017), and model parameters were estimated under an edge-linked partition scheme with a free evolutionary rate, and each orthogroup treated as a separate partition. A maximum-likelihood (ML) tree was then produced in IQTree using the edge-linked partition model, and branch support was assessed with 1,000 ultrafast bootstrap replicates (UFBoot) (Hoang et al. 2018; Nguyen et al. 2015). IQTree was also used to infer the 2,214 individual ML gene trees, again using ModelFinder. A coalescent-based species tree was reconstructed with the individual orthogroup trees using ASTRAL-III v5.7.1 (Zhang et al. 2018). Bipartition support was assessed via local posterior probabilities (LPP, '-t 3' option in ASTRAL; (Sayyari et al. 2016)). Genealogical

concordance was assessed using gene concordance factors (gCF) and site concordance factors (sCF) calculated in IQTree (Minh et al. 2020).

Secondary metabolite gene cluster analysis

Secondary metabolite biosynthetic gene clusters (BGCs) were predicted from the mycobiont proteomes using the command line implementation of antiSMASH v5.1.2 (Blin et al. 2019). Homology between gene clusters was inferred by clustering antiSMASH BGCs into Biosynthetic Gene Cluster Families (BGCFs) using BiG-SCAPE (Navarro-Muñoz et al. 2020). BiG-SCAPE calculates pairwise distances between BGCs using a combination of protein domain content, sequence identity, domain order and copy number information. It then links all pairs of BGCs with a distance score below a set cut-off into a sequence similarity network and performs affinity clustering on each group of connected nodes to assign BGCs to families. I used a range of BiG-SCAPE cut-off values from 0.3 to 1.0 (default to maximum) in my comparisons. BGCs from the MIBiG database v1.4 (Medema et al. 2015) were included in the BiG-SCAPE analysis to link my mycobiont BGCs to published BGCs with known metabolite products. Following Theobald et al. (2018), I chose the minimum cut-off that ensured the MIBiG clusters for the highly similar compounds TAN-1612 and asperthecin were grouped into the same family. This led to a final cut-off of 0.46. BiG-SCAPE was run with the '-hybrids-off' option to ensure that each BGC only appeared in one family. Ribosomally synthesized and post-translationally modified peptide (RiPPs) and saccharide clusters were excluded to reduce computational time.

I estimated turnover in BGC content between the genomes using beta diversity (Baselga 2010), which in this context represents the extent to which dissimilarity in BGC content is due to each genome containing unique sets of BGCs rather than the less diverse genomes simply being subsets of the more diverse ones. I observed whether BGCF presence-absence patterns reflected phylogenetic relatedness between taxa using principal coordinates analysis (PCoA) and then tested this using a Mantel test and permutational analysis of variance (PERMANOVA). Firstly, I converted the ML tree into a phylogenetic distance matrix using the cophenetic.phylo function of the R package ape v5.6-1 (Paradis et al. 2019). Subsequently, the BiG-SCAPE BGCF presence-absence matrix (excluding non-lichen MIBiG clusters) was transformed into a dissimilarity matrix by calculating Jaccard dissimilarity indices using the 'vegdist' function of the R package vegan v.2.5-7. The BGCF dissimilar-

ity matrix was used as input for the PcoA using vegan's 'wcmdscale' function. Correlation between the phylogenetic and BGCF distance matrices was tested using a Mantel test implemented with vegan. For PERMANOVA I followed the approach of Mesny et al. (2020); to summarise, a principal component analysis (PCA) of phylogenetic distances was performed using the function PCA from the Python package scikit-learn v0.21.3 (Python v3.7.4). The first two principal components (PC1 and 2) were sufficient to cover over 80% of the variance due to phylogeny and were used alongside the Jaccard BGCF distance matrix for PERMANOVA analysis using the adonis2 function of vegan with the model BGCF_DistanceMatrix ~ phylogeneticPC1 + phylogeneticPC2 (Mesny et al. 2020).

Putative anthraquinone gene clusters were identified by searching for families that contained experimentally identified anthraquinone clusters from MIBiG. For families containing these MIBiG clusters, I further investigated PKS evolution by extracting the core PKS enzyme gene from each BGC. I identified which orthogroups these proteins belonged to and pulled all the amino acid sequences belonging to that orthogroup. Amino acid sequences were aligned using MAFFT, and poorly aligned sequences were detected using GUIDANCE2 (Sela et al. 2015). These sequences were removed from the original multifasta file, and amino acids were aligned again. The multiple sequence alignment was converted into nucleotides using PAL2NAL (Suyama et al. 2006), and a maximum-likelihood (ML) gene tree was then reconstructed using RAxML-ng with a GTR+G substitution model (Kozlov et al. 2019). Branch support was assessed with Felsenstein's bootstraps and the Transfer Bootstrap Expectation metric (Lemoine et al. 2018). Bootstrapping convergence was tested using the autoMRE criterion within RAxML-ng.

I then performed functional annotation on the amino acids of the PKS orthogroup. I classified proteins using Pfam domains in the InterPro database (Blum et al. 2021) via InterProscan (`https://github.com/ebi-pf-team/interproscan`). InterProscan also includes PANTHER analysis that suggests protein function. In addition to the core PKS enzyme genes, I searched for other genes potentially shared among all my putative anthraquinone clusters but absent from non-anthraquinone BGCs. These additional genes were functionally annotated in the same way.

Finally, I analysed the product template (PT) domains of the core PKSs to provide further evidence as to whether my putative anthraquinone clusters could synthesise anthraquinones. The PT domain is essential for cyclisation of the final polyke-

tide and is a key determiner of the final metabolite structure (Crawford et al. 2009). I extracted the product template (PT) domain from my PKSs with BLASTp (Camacho et al. 2009), using the asperthecin AptA PT domain from *Aspergillus nidulans* as a query (Szewczyk et al. 2008). My lichen PT domains were combined with PT domains from a curated dataset of fungal non-reducing PKSs (NR-PKSs) with confirmed metabolite products from (Liu et al. 2015a), which were downloaded from NCBI. The PT domains were then aligned to the Pfam seed alignment for the polyketide PT domain (PS-DH PF14765: https://pfam.xfam.org/family/PF14765) using a Hidden Markov Model (HMM) alignment via the following perl script: `https://github.com/reubwn/scripts/blob/master/hmmsearch-easy.pl`. Pfam seed alignments reflect the diversity of a particular protein domain across the tree of life and thus help to place my potential anthraquinone genes within a wider evolutionary context. I then reconstructed a maximum likelihood tree from the resulting PT domain alignment using IQTree with 1,000 ultrafast bootstraps and ModelFinder to determine the optimal substitution model. All bioinformatics scripts for my pipeline can be found here: `https://github.com/theo-llewellyn/TeloschistalesMetagenomics`.

2.4 Results

Metagenomics produces new high-quality mycobiont genomes

I present 24 newly sequenced and de novo assembled *Lecanoromycetes* mycobiont genomes from metagenomic data, covering four families in three orders: the *Letrouitiaceae* and *Teloschistaceae* (*Teloschistales*), the *Umbilicariaceae* (*Umbilicariales*), and the *Graphidaceae* (*Ostropales*). Due to the high level of taxonomic uncertainty and ongoing revision of *Teloschistaceae* genus names, I opted for a conservative approach and used only established names as in Arup et al. (2013) (for example *Teloschistes* over *Niorma*). The metagenome assembler MEGAHIT (Li et al. 2015) consistently produced the most complete (BUSCO %) and contiguous (N50) assemblies compared to MetaSPAdes (Nurk et al. 2017), which only produced higher quality assemblies in a handful of cases (taxa labelled with S in Table 2.1).

I developed a new pipeline to isolate mycobiont reads from the lichen metagenome using an iterative, combined approach of the Blobtools workflow (Laetsch et al. 2017),

Table 2.1: Genome assembly statistics for the assembled *Lecanoromycetes* mycobiont genomes. Taxa with superscript S highlight assemblies where MetaSPAdes performed best. Assembly size is indicated in megabase pairs and N50 in thousands of bases.

Taxon	Assembly size (Mbp)	Scaffolds	N50 (K)	BUSCO (%)	Dup. BUSCO %	Genes	Accession
Caloplaca aegea	31.48	6484	9	92.44	0.05	7543	JALAIP 000000000
C. aetnensis	31.90	2773	24	95.72	0.06	9101	JALAIC 000000000
C. ligustica	24.20	1915	19	89.27	0.23	7863	JALAIE 000000000
Diploschistes diacapsis	29.15	4361	15	91.38	0.00	8138	JALAIB 000000000
Flavoplaca oasis	33.60	3025	28	95.13	0.17	9027	JALAIS 000000000
Gyalolechia ehrenbergii^{S}	39.67	317	209	97.25	0.00	8095	JALAIJ 000000000
Letrouitia leprolyta^{S}	27.85	1220	34	94.37	0.11	9953	JALAIH 000000000
L. transgressa 1^{S}	26.34	337	290	95.90	0.29	8199	JALAIO 000000000
L. transgressa (2)	27.99	501	211	96.37	0.23	8581	JALAIG 000000000
Seirophora lacunosa 1^{S}	38.78	428	146	97.89	0.06	8161	JALAHX 000000000
S. lacunosa (2)	35.69	2533	21	94.49	0.06	7847	JALAIF 000000000
S. villosa^{S}	36.72	2163	28	93.73	0.11	7507	JALAHY 000000000
T. chrysophthalmus	27.03	776	106	97.95	0.35	8751	JALAIL 000000000
Teloschistes flavicans	25.62	1520	25	94.49	0.06	8438	JALAIQ 000000000
T. peruensis	26.73	966	82	97.71	0.41	8483	JALAHW 000000000
Umbilicaria vellea	28.68	3107	14	90.56	0.17	7140	JALAID 000000000
Usnochroma carphinea^{S}	36.98	2385	30	97.36	0.23	8370	JALAHV 000000000
Variospora aurantia	24.80	5327	8	89.92	0.06	7864	JALAHZ 000000000
Xanthomendoza fulva	29.80	952	129	97.83	0.29	8822	JALAIN 000000000
Xanthoria aureola^{S}	38.52	5150	12	96.54	0.53	10011	JALAII 000000000
X. mediterranea^{S}	32.48	447	140	97.89	0.12	9057	JALAIA 000000000
Xanthoria sp. 1	54.66	8581	11	90.68	1.14	11606	JALAIM 000000000
Xanthoria sp. 2	37.07	2884	27	94.43	0.17	9784	JALAIR 000000000
X. steineri^{S}	30.64	2267	35	97.71	0.23	9436	JALAIK 00000000

UniRef and custom lichen database blast searches, and CONCOCT metagenome binning (Alneberg et al. 2014) (see Supplementary Figure A.1 for an example of the pipeline output). I developed this pipeline as exploratory data analysis showed AT-rich regions

that make up significant proportions of the mycobiont genome were not being assigned to the mycobiont using BLAST or database methods like Kraken (Wood et al. 2014), an issue also observed by Tagirdzhanova et al. (2021). The full pipeline with accompanying bioinformatics scripts is available here: `https://github.com/theo-llewellyn/TeloschistalesMetagenomics`.

Mycobiont assembly sizes range from 25.18 megabase pairs (Mbp) (*Caloplaca ligustica*) to 54.66 Mbp (*Xanthoria* sp. 1) with a mean assembly size of 32.47 Mbp. These values fall within the range of published *Lecanoromycetes* genome assemblies sequenced from cultured samples (Supplementary Figure A.2). All assemblies were over 90% complete for BUSCO genes (mean 94.88%) except for *Caloplaca ligustica* (89.27%) and *Variospora aurantia* (89.92%). Duplicated BUSCO percentages were all below 0.6% (mean 0.21%) except for *Xanthoria* sp. 1 (1.14%). Completeness scores are comparable to, or higher than, published *Lecanoromycetes* genomes obtained from cultures. The number of predicted genes varied considerably, ranging from 7,140 (*Umbilicaria vellea*) to 11,606 (*Xanthoria* sp. 1).

Assembly quality metrics were comparable to those from published genomes obtained from cultured mycobionts, yet with slightly more variation (Supplementary Figure A.2). Metagenome assemblies were generally more fragmented than genomes sequenced from cultured samples (Supplementary Figure A.2), an unsurprising result given that high molecular weight DNA is easier to extract from an actively growing culture. Assembly fragmentation does not appear to have affected gene annotation, with a similar number of predicted genes between culture-derived and metagenomic assemblies (Supplementary Figure A.2).

I used both concatenation and coalescent-based approaches to reconstruct a whole-genome scale phylogeny for my 24 newly sequenced mycobiont genomes, combined with a further 21 published *Lecanoromycetes* genomes. The final alignment consisted of 2,214 single copy protein-coding genes. IQTree (Nguyen et al. 2015) and ASTRAL-III (Zhang et al. 2018) recovered largely the same topology except for two branches within the family *Parmeliaceae* (order *Lecanorales*) (red branches in Figure 2.1). Differing branches showed high gene and site conflict, estimated using gene and site concordance factors (Minh et al. 2020). All internodes were significantly supported (UFBoot and LPP >95%) except for the monophyly of the genus *Umbilicaria* and the conflicting relationship within the

Parmeliaceae. All families were recovered as monophyletic with >80% gene and >50% site concordance, suggesting low gene and site conflict. The three subfamilies identified in the *Teloschistaceae* (Gaya et al. 2012; Arup et al. 2013) were supported here, and all *Teloschistales* genera were monophyletic. The presence of a sample of *Letrouitia leprolyta* within *Gyalolechia* confirms a suspected erroneous microscopic identification and is likely to be described as a new taxon. Further study of this sample will be required.

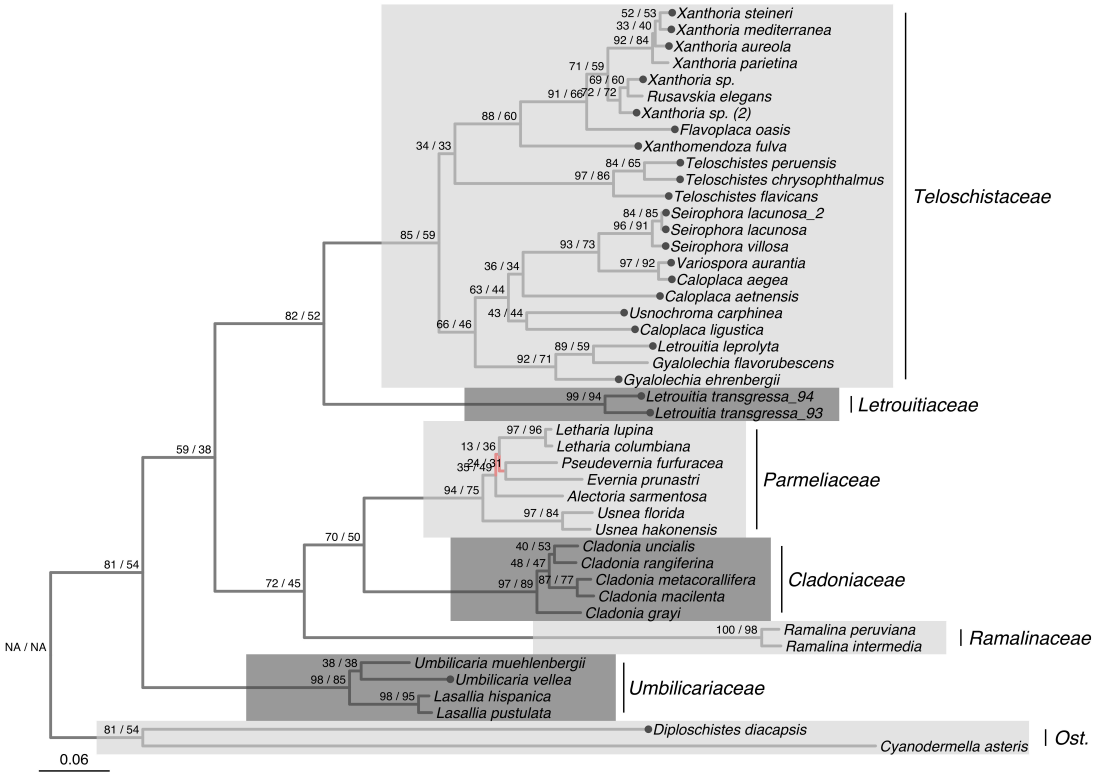


Figure 2.1: Maximum Likelihood tree including representatives of the largest class of lichenised fungi, the *Lecanoromycetes* (*Ascomycota*), with special focus on the order *Teloschistales*, reconstructed with IQTree using a 2,214 single-copy gene combined dataset. All internodes, except for the one connecting *Umbilicaria muehlenbergii* and *U. vellea* and the red branches in the family *Parmeliaceae*, were significantly supported by both bootstrap and posterior probabilities. Red branches indicate a conflict with the ASTRAL tree. Support values before and after the forward slash indicate gene and site concordance factors, respectively. Dotted tips show taxa sequenced and assembled in this study. Families/Order are labelled and highlighted by alternating dark and light and grey blocks. Ost. = *Ostropales*.

Lecanoromycetes genomes are diverse in secondary metabolite gene clusters and show high dissimilarity

Annotation of biosynthetic gene clusters (BGCs) in the 45 *Lecanoromycetes* genomes (24 de novo assemblies and 21 previously published) identified a total of 2,073 BGCs. The average number of BGCs per genome was 46 ±12.80. In a fragmented assembly, large BGCs may be split across multiple contigs, which could lead to them being erroneously detected

as two or more separate clusters. I used a linear model implemented in R to test whether assembly fragmentation (N50) affected number of BGCs detected per genome. A partial F-test showed no significant effect of assembly fragmentation (N50) on number of detected BGCs (Supplementary Figure A.3). Therefore, it does not appear that I overestimated the number of BGCs per genome in this study. Type 1 polyketide synthases (PKSI) and non-ribosomal peptide synthases (NRPS) were the most abundant BGC classes across all genomes (Figure 2.2c). BiG-SCAPE network analysis (Navarro-Muñoz et al. 2020) grouped the BGCs into 1,033 biosynthetic gene cluster families (BGCFs) using pairwise distances calculated from similarity of BGC protein domain content, sequence identity, domain order and copy number. I included reference BGCs with known metabolite products from the MIBiG database (Medema et al. 2015) to infer putative BGC metabolite products in my lichen BGCs. Of the 1,033 BGCFs, 22 contained MIBiG reference BGCs (Table 2.2). The largest BGCF was an NRPS family containing 25 BGCs. The majority of the 22 families containing MIBiG reference BGCs were PKSIs.

Table 2.2: Results of BiG-SCAPE clustering of BGCs into BGC families.

BGC class	**No. families**	**Average no. BGC/family**	**Largest family (No. BGCs)**	**Families w/ MIBiG BGC**
PKSI	421	2	16	16
NRPS	361	2	25	0
PKS-NRP Hybrid	73	2	13	4
PKS other	21	3	16	1
Terpene	99	2	24	1
Others	58	2	5	0
TOTAL	1033	2	NA	22

BGC families mostly contained BGCs present in one or two taxa/genomes (Figure 2.2d top plot), suggesting that BGCs are highly dissimilar between taxa. I estimated BGC content turnover between genomes using beta diversity, which in this context represents the proportion of BGC dissimilarity due to each genome containing a unique set of BGCs rather than a case of nestedness where genomes with fewer BGCs are simply subsets of more diverse genomes (Baselga 2010). Beta diversity (Sorensen's) showed BGC content was highly dissimilar between genomes and this was due mainly to turnover (Sorensen's beta diversity = 0.976: turnover=0.971, nestedness=0.005), suggesting each genome contains a largely unique set of BGCs. To test whether the high dissimilarity of BGCs was due to BiG-SCAPE being too strict when grouping BGCs into families, I performed the

same analysis at the most relaxed cut-off value the algorithm permitted. The relaxed cut-off results showed a lower proportion of singleton BGCFs; however, the overall pattern remained with the majority of BGCFs being represented by less than five taxa (Figure 2.2d bottom). Beta-diversity decreased slightly to 0.959, with the majority still due to turnover (turnover=0.95, nestedness=0.009).

Some blocks of BGCFs in Figure 2.2b (showing only PKSI class for readability) appear to show phylogenetic patterns to their distributions. A principal coordinates analysis of BGCF profile dissimilarity (measured via Jaccard dissimilarity indices) suggests some heritability in BGCF profiles, as colouring the points by taxonomic order reveals four groups of data points representing the orders *Lecanorales*, *Ostropales*, *Teloschistales*, and *Umbilicariales* (Figure 2.2e), with little overlap between the groups. A Mantel test showed a significant correlation between the phylogenetic distance of taxa and their BGCF profile dissimilarity as measured via Jaccard dissimilarity indices (Mantel statistic R: 0.6696, p-value = 0.001). PERMANOVA analysis further supported this, showing phylogenetic signal significantly explained variation in BGCF profile dissimilarity albeit weakly (phylogenetic principal component 1: $R^2 = 0.06998$, p-value = 0.0001, phylogenetic principal component 2: $R^2 = 0.03987$, p-value = 0.0001).

Gene cluster family analysis identifies putative anthraquinone clusters in *Lecanoromycetes* genomes

To identify putative anthraquinone BGCs, I restricted my search to PKSI BGCFs since anthraquinones are aromatic polyketides. I searched for families containing reference MIBiG clusters linked to anthraquinones or structurally similar compounds. Four such families (BGCFs 45, 3197, 3070 and 2157) were found, containing 39 *Lecanoromycetes* BGCs linked to following the MIBiG clusters: the anthraquinones emodin in *Escovopsis weberi*, asperthecin in *Aspergillus nidulans* FGSC A4 and endocrocin/clavorubin in *Claviceps purpurea*, and the structurally related alternariol in *Aspergillus nidulans* FGSC A4 and TAN-1612 in *Aspergillus niger*. The BiG-SCAPE similarity network connected all four BGCFs into a single network, further suggesting that gene clusters within these BGCFs are involved in the biosynthesis of similar compounds (Figure 2.3a). The distribution of these four BGCFs across the *Lecanoromycetes* genome-scale phylogeny is shown in Figure 2.3b. All four BGCFs are present in the *Teloschistales*, and all *Teloschistales* genomes contain

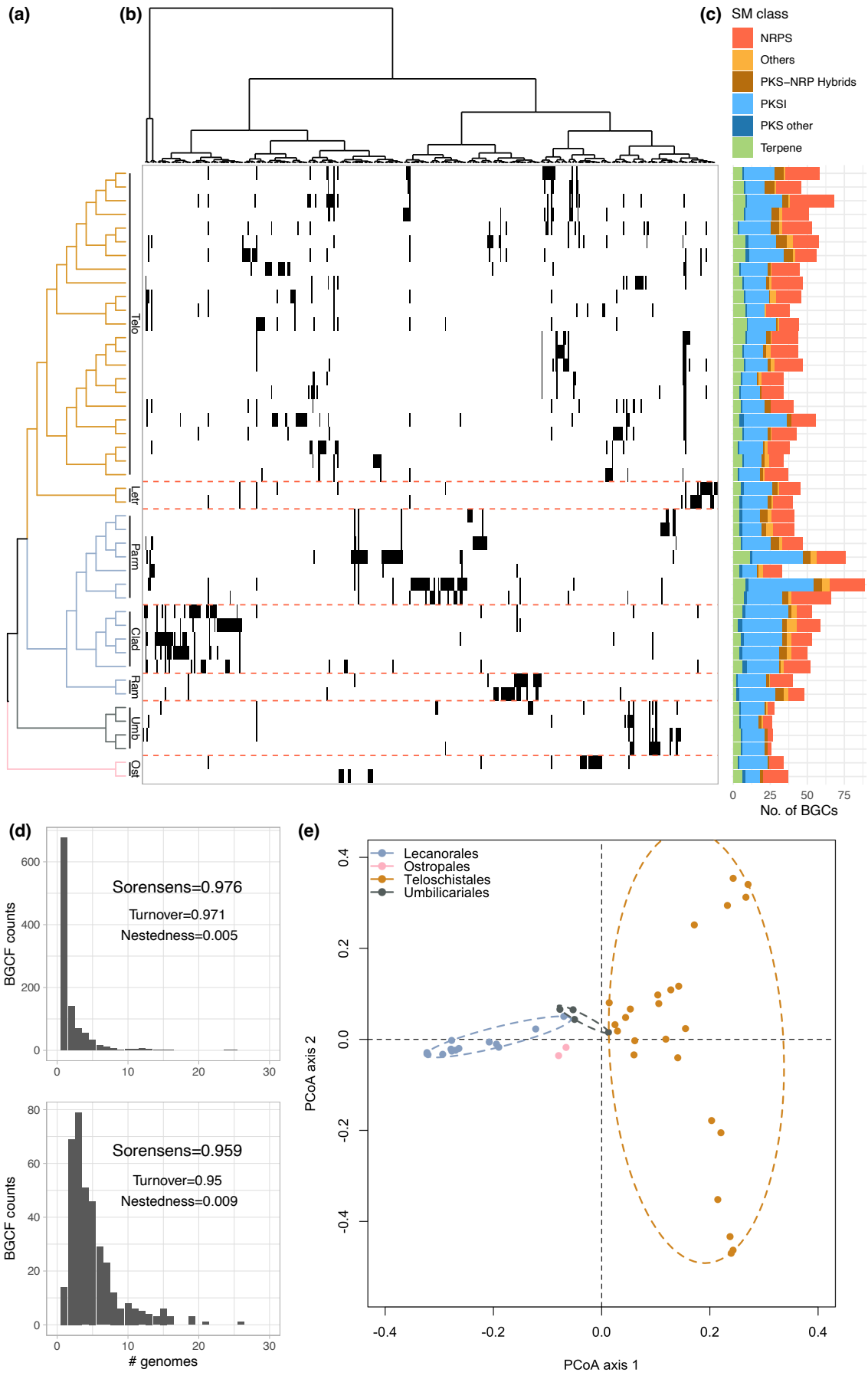


Figure 2.2: (Caption next page.)

Figure 2.2: Diversity and distribution of secondary metabolite biosynthetic gene clusters (BGCs) in the *Lecanoromycetes*. (**a**) Topology represents the ML tree shown in Figure 2.1 with branches coloured by taxonomic order as per Figure 2.2e legend. (**b**) Presence-absence matrix of BGCFs in each genome (rows), showing only PKSI class BGCFs for readability. Each matrix column represents a single BGCF, with columns ordered by presence-absence similarity profiles across genomes. Dendrogram above matrix displays the BGCF similarity profile clustering produced with hclust function in R. Horizontal, orange dashes delineate taxonomic families. Presence = black, absence = white. (**c**) Stacked bar plot showing the number of predicted BGCs per genome. Colours represent different BGC classes. (**d**) Histograms showing the number of genomes per biosynthetic gene cluster family (BGCF). The top histogram shows the BiG-SCAPE results using a strict cut-off of 0.46 to group BGCs into BCGFs and the bottom histogram shows the results using a more relaxed cut-off of 1. Beta diversity statistics at both cut-offs are shown on each graph and are separated into their turnover and nestedness components. (**e**) Principal coordinates analysis (PCoA) of all BGCF presence-absence profiles shown in (**b**). Colours represent taxonomic orders with ellipses drawn around orders with 3 or more points.

a gene cluster in at least one of the four BGCFs. BGCF 3197 is also found in the *Umbilicariaceae* and *Cladoniaceae*, and BGCF 45 in the *Umbilicariaceae* and *Parmeliaceae*.

I then extracted the core PKS amino acid sequence from the 39 *Lecanoromycetes* BGCs belonging to these four BGCFs. All 39 PKS amino acid sequences belong to a single group of homologous genes (orthogroup), containing a further 301 sequences that were not present in my four putative anthraquinone BGCFs. The phylogenetic gene tree of this orthogroup is shown in Figure 2.3c (amino acids were aligned and filtered using MAFFT and GUIDANCE2, and tree reconstructed using RAxML-ng, see methods for full details). The putative anthraquinone PKS sequences fall within a narrower supported monophyletic group within the gene tree (grey box in inset and zoomed in clade in Figure 2.3c) that contains an additional 17 sequences indicated with grey dots at tips. Therefore, the core PKS genes in my putative anthraquinone BGCs are homologous to many other PKS genes, including PKSs from lichens that are not known to produce anthraquinones.

Anthraquinone BGCs in the *Teloschistales* clade have a conserved four-gene structure including a unique ABC transporter

Functional annotation with PANTHER suggested three potential functions for the PKS sequences within my putative anthraquinone PKS orthogroup (Supplementary Figure A.4a). When plotting these functions on the orthogroup gene tree (Supplementary Figure A.4b), a distinction is visible with most proteins being annotated as conidial pigment polyketide synthase ALB1 proteins, whilst my putative anthraquinone PKSs (shown in Figure 2.3c) were annotated as either 'atrochrysone carboxylic acid synthase-related' or AptA-related. Atrochrysone carboxylic acid is a common intermediate in anthraquinone biosynthesis in non-lichenised *Ascomycota* and AptA is the *Aspergillus nidulans* gene involved in the

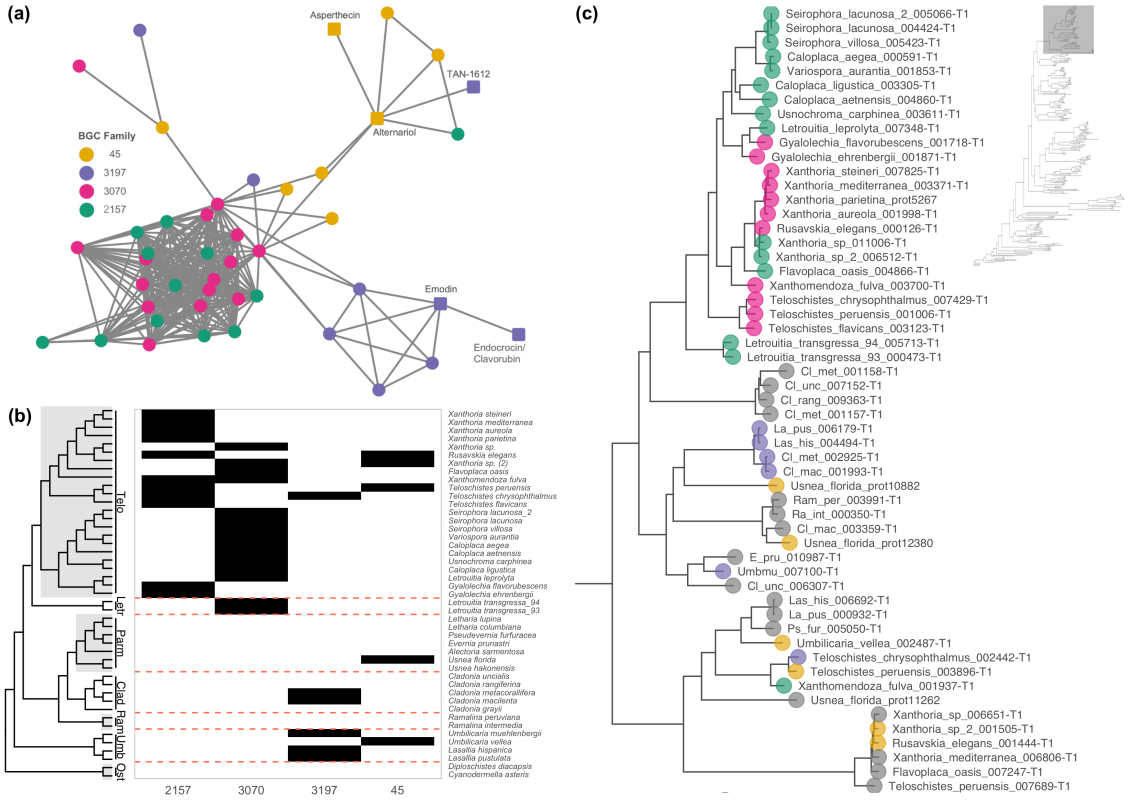


Figure 2.3: Putative anthraquinone Biosynthetic Gene Cluster (BGC) families. (**a**) BiG-SCAPE similarity network linking BGCs (nodes) that showed a pairwise similarity score below the 0.46 cut-off established in this study. Node colour shows to which BGC family each BGC was assigned by affinity clustering. Square nodes indicate MIBiG reference gene clusters that have been experimentally linked to a compound via genetic manipulation. (**b**) Presence-absence matrix of the four putative anthraquinone BGC families plotted against the phylogenomic tree shown in Figure 2.1. Orange dashed lines delineate lichen taxonomic families. (**c**) Zoomed-in clade of the PKS maximum-likelihood gene tree showing the 39 core PKS genes for the BGC families shown in (**a**) and (**b**). Coloured dots at tips show which BGC family the 39 core PKS genes belong to, grey dots are sequences that do not belong to the four putative anthraquinone BGCFs. Full tree, including all 340 sequences found for this orthogroup, is depicted in the inset, where the clade of interest is highlighted with a grey rectangle.

production of the anthraquinone asperthecin.

I followed PANTHER analysis with protein domain analysis of my putative anthraquinone PKS amino acid sequences. This identified four of the five key Pfam domains expected within a non-reducing PKS protein: SAT-KS-AT-ACP. Whilst the canonical product template domain (PT), the fifth key PKS domain, was not detected using Pfam, subsequent BLAST analysis showed it to be present in all *Teloschistales* anthraquinone PKSs. Pfam also failed to annotate the PT domain of the MIBIG anthraquinone BGCs linked to my putative anthraquinone BGCs. This suggests the PT domains in fungal anthraquinone clusters may contain mutations that prevent them from being annotated as such by Pfam.

In the putative anthraquinone BGCs belonging to *Teloschistales* taxa, besides the core

PKS gene, three additional genes were consistently and uniquely present (Figure 2.4a). The first gene encodes an ATP-Binding Cassette (ABC) protein, a transmembrane efflux pump commonly found in fungal cells that actively expel metabolites. The second gene was annotated as encoding a metallo-β-lactamase type-thioesterase (MβL-TE). Thioesterases cleave the final polyketide from the PKS enzyme and can exist either as a domain within a PKS protein or as a separate protein. The third additional gene was annotated as encoding a dehydratase with a single EthylD domain. For *Usnochroma carphinea*, this domain was instead found within the PKS gene. When present, the order of these four genes was the same across all my putative anthraquinone gene clusters in the *Teloschistales*, although some clusters contained additional genes. Some *Teloschistales* genomes showed exceptions to this four-gene core structure, however. In addition to the aforementioned *Usnochroma carphinea* variation, *Seirophora villosa* was missing the EthD gene, *Xanthoria sp.* 1 was missing the ABC transporter, and *Caloplaca aegaea* only had the core PKS gene. Whilst non-*Teloschistales* BGCs in my putative anthraquinone group all contained the PKS gene, and some also contained the EthD and/or MβL-TE genes, the ABC transporter gene was exclusive to *Teloschistales* BGCs.

Phylogenetics shows putative anthraquinone PKSs from lichenised *Ascomycota* are highly syntenic with those from non-lichenised *Ascomycota*

Given that the PT domain is indicative of compound structure and has been used to classify fungal PKSs, I extracted the PT domains from my 39 anthraquinone PKSs and combined them with 44 additional PT domains from fungal NRPKSs with known metabolite products (curated by Liu et al. (2015a)) for phylogenetic analysis. I aligned these PT domains with a further 76 bacterial and fungal PT domains from the Pfam seed alignment (PF14765) using a hidden Markov model approach (see methods). Eight sequences did not pass the filtering/quality control steps of the script leading to a final alignment of 151 sequences.

In the reconstructed maximum likelihood tree, my putative anthraquinone PT domains formed a significantly supported grouping with the Liu et al. (2015a) fungal PT domains with known products (Figure 2.4b). Within this clade, five supported subgroups (posterior probability >0.95) are highlighted in Figure 2.4b. As these subgroups correspond exactly to five of the eight groups identified by Liu et al. (2015a), I labelled them using the

same numbering. My putative anthraquinone PT domains (orange names at tips in Figure 2.4b) fell within group V along with sequences from non-lichenised fungal PKSs responsible for the anthraquinones endocrocin, atrochrysone/emodin and atrochrysone carboxylic acid/monodictyphenone, and the anthraquinone-related compounds alternariol and TAN-1612 (green tip names in Figure 2.4b). Within group V all *Teloschistales* anthraquinone PT domains form a single supported group. *Teloschistes peruensis* and *Xanthomendoza fulva* have an additional copy of the PKS gene (sequences Teloschistes_peruensis_003896 and Xanthomendoza_fulva_001937 in Figure 2.4b), which falls outside this supported group but still within group V.

2.5 Discussion

This study presents 24 new lichenised fungal genomes, focusing on the diverse and undersampled *Teloschistales* order. My metagenomic approach successfully isolated mycobiont genomes from lichen metagenomes, producing complete assemblies with good contiguity. A whole-genome phylogeny combining these new genomes with 21 other published genomes (20 lichenised and one non-lichenised fungus) of the *Lecanoromycetes* class shows high support and low conflict for all families and all but one genera included. Secondary metabolite biosynthetic gene clusters (BGC) are numerous in the studied lichen genomes, and comparative analysis shows BGC repertoires are highly dissimilar among taxa. Comparative genomics revealed a set of BGCs potentially responsible for anthraquinone biosynthesis. Functional annotation of these clusters supported their involvement in anthraquinone metabolism and uncovered a conserved four-gene BGC structure I believe to be shared uniquely across the *Teloschistales*. Further phylogenetic analysis of anthraquinone PKS product template (PT) domains, a key domain in classifying fungal PKS function, showed *Teloschistales* PT domains were highly similar to those of anthraquinone PKSs from non-lichenised fungi, providing additional support for them producing anthraquinones.

Metagenomics is a reliable alternative to axenic cultures but still requires manual data curation

Metagenome-assembled genomes (MAGs) are now commonplace in the field of lichen genomics and have consistently been demonstrated as reliable alternatives to culture-derived

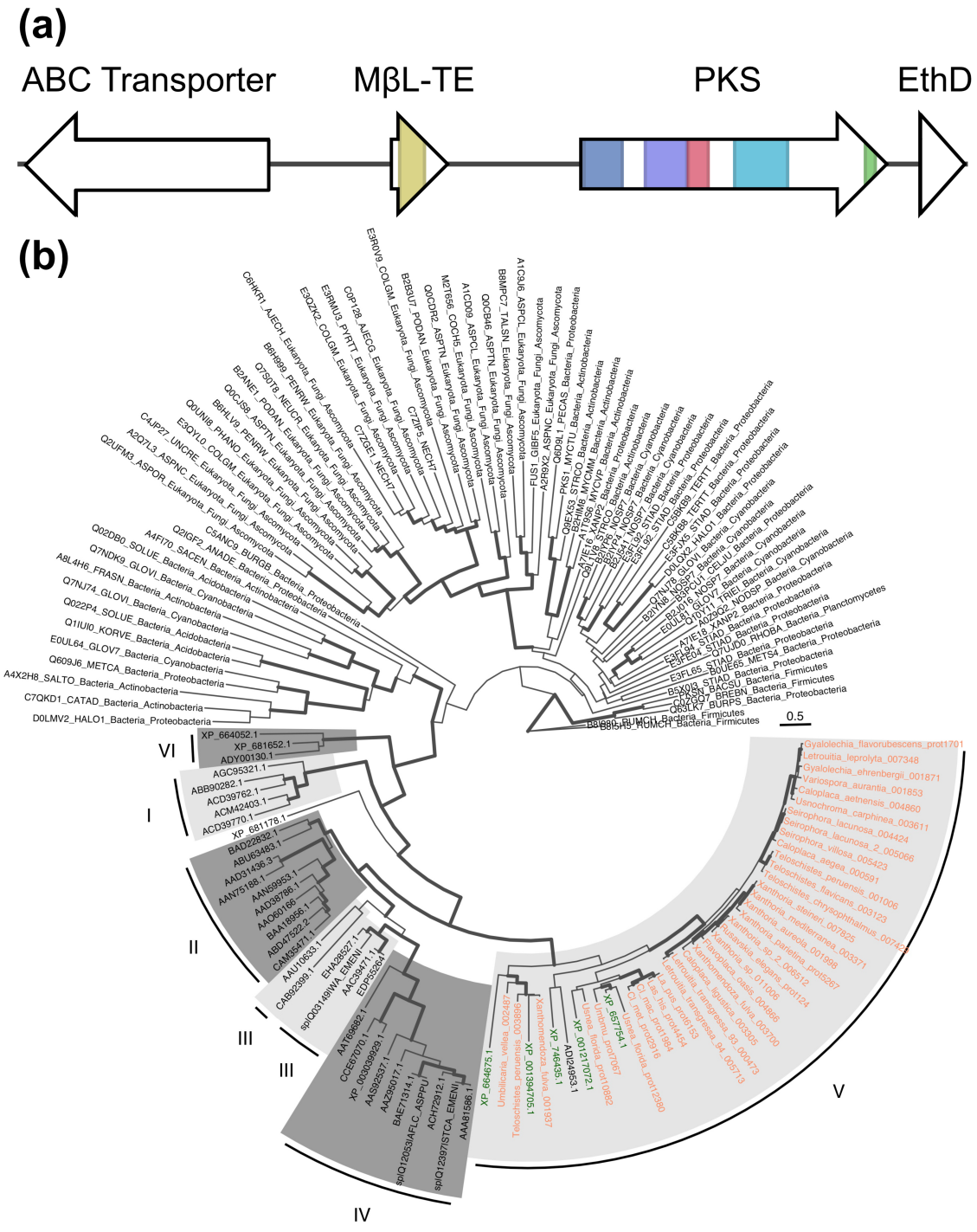


Figure 2.4: (**a**) Biosynthetic gene cluster (BGC) showing the four common genes in *Teloschistales* putative anthraquinone BGCs. From left to right: ATP-Binding cassette (ABC) transporter, metallo-β-lactamase type thioesterase (MβL-TE), PKS (with SAT-KS-AT-ACP domains), and dehydratase gene containing EthD domain. BGC shown is from the *Variospora aurantia* genome. (**b**) Maximum-Likelihood phylogeny reconstructed using the Product template (PT) domain (Pfam: PF14765) and a hidden-Markov model alignment. Bold branches indicate bootstrap support ≥95%. Alternating dark and light grey blocks highlight the PT classification of Liu et al. (2015a). All lichenised-fungal putative anthraquinone sequences from this study fall within group V. Putative *Lecanoromycetes* anthraquinone sequences are those with orange text and sequence ADI24953.1. Green text signifies sequences from genes experimentally confirmed to produce anthraquinones.

genomes (Meiser et al. 2017; Greshake Tzovaras et al. 2020; McKenzie et al. 2020; Singh et al. 2021b; Singh et al. 2022). The bioinformatic analysis of lichen metagenomes still presents challenges, however. For example, in this study, the mycobiont filtering step, though effective, required manual curation when deciding which metagenome bins to merge. Tools such as metaWRAP (Uritskiy et al. 2018) can automate this process but are currently restricted to prokaryote datasets (Parks et al. 2015). Increasingly, scientists are recognising the importance of eukaryotes (especially fungi) in microbial communities (West et al. 2018) and eukaryote-focused tools are being developed (Levy Karin et al. 2020; Saary et al. 2020). Euk-CC (Saary et al. 2020) is one example that has recently been tested on a lichen metagenome (Tagirdzhanova et al. 2021), yet still required manual merging of bins to capture the AT-rich regions of the mycobiont genome.

AT-rich regions have been frequently observed in non-lichenised fungal genomes, often constituting a significant proportion of the total genome size and containing genes with important adaptive roles (Rouxel et al. 2011; Testa et al. 2016). Based on my results, it seems these regions are also relevant for lichen mycobionts and should not be excluded during mycobiont filtering. Following Tagirdzhanova et al. (2021), I tested Euk-CC on my dataset but found that it failed to recover significant proportions of the mycobiont genomes. Therefore, I developed my own pipeline combining CONCOCT and BlobTools with UniRef and custom lichen database blast searches, which worked successfully on lichens across diverse lineages. Going forward, new lichen mycobiont genomes can be used as training data to develop lichen-specific metagenome binning tools and address this analytical bottleneck.

In addition, we still face shortcomings in assessing 'completeness' of fungal genome assemblies using current bioinformatic tools. Assembly size can differ significantly from genome sizes estimated via experimental approaches such as cytometric analysis. Even highly complete BUSCO scores, which quantify the presence of a set of lineage-specific core genes, can be seen in genomes that are less than 50% complete based on cytometric data (Le Cam et al. 2019; Hill et al. 2021). Flow cytometry would provide an important benchmark when assessing the effectiveness of my novel metagenomic bioinformatic pipeline to isolate complete mycobiont genomes. Unfortunately, cytometric data of lichenised fungi is still lacking (Hill et al. 2021), again due to difficulties with culturing lichen mycobionts. However, axenic cultures are not always necessary for cytometric analysis, as demonstrated

by the work of Tavares et al. (2014) on fungal plant pathogens.

My results support the idea that metagenomics can be used to target under-represented lineages in the *Lecanoromycetes*, such as the *Teloschistales*. Recently, Resl et al. (2022) implemented a similar approach to generate 18 MAGs covering diverse lineages and lifestyles of the *Lecanoromycetes*. Lichen MAGs can therefore be successfully incorporated into whole-genome scale phylogenetic trees to investigate evolutionary and systematics-based hypotheses (Resl et al. 2022). The phylogenomic tree reconstructed in this study (Figure 2.1) agrees with other genome-scale trees of the *Ascomycota* (Shen et al. 2020; Li et al. 2021). My tree is also largely in agreement with Pizarro et al. (2018)'s genome-scale tree of the *Parmeliaceae* family, except for the placement of *Usnea*. Pizarro et al. (2018) recover *Usnea* as sister to the *Evernia*/*Pseudevernia* clade, whereas I recover it as sister to the rest of the *Parmeliaceae* albeit with low gene and site concordance. At the order level, the phylogenomic tree shown in Figure 2.1 matches that of the most comprehensively sampled and complete multigene phylogeny of the *Lecanoromycetes* (Miadlikowska et al. 2014).

Lichen PKS clusters are highly dissimilar but share core PKS genes

The increased number of sequenced fungal genomes allows for comparative analysis of BGCs amongst many species. Here I performed a comparative investigation of entire BGC profiles in lichens and the *Lecanoromycetes* class in particular. I observed high dissimilarity of BGC profiles between *Lecanoromycetes* genomes. This is not the first time that such high biosynthetic gene cluster diversity has been observed in fungi. In non-lichenised *Ascomycota*, similar patterns of dissimilarity have been observed at phylum (Robey et al. 2021), class (Gluck-Thaler et al. 2020), genus (Vesth et al. 2018), and species level (Lind et al. 2017). In lichens, BGC content can even vary between populations of the same species at different altitudes (Singh et al. 2021a).

The absence of BGCs in certain clades may be explained by missing genome regions due to incomplete assemblies. However, I deem this unlikely as gene completeness scores were high across all analysed genomes, and the same pattern of high BGC dissimilarity has been observed in studies that incorporate long-read, reference-quality assemblies (e.g. (Gluck-Thaler et al. 2020)). Previous work has linked the high dissimilarity of fungal BGCs to speciation (Gluck-Thaler et al. 2020; Lind et al. 2017). Since the lichen clades in this study cover distant taxonomic groups and diverse ecological niches, speciation

processes such as niche differentiation may well be contributing to the high dissimilarity I observe in the *Lecanoromycetes* BGC profiles, as shown by Singh et al. (2021a) at the population level. Despite the high BGC dissimilarity, Mantel and PERMANOVA tests both identified significant effects of phylogenetic relatedness on BGC profile dissimilarity, suggesting that closely related taxa still have more similar BGC profiles than distantly related taxa.

I aimed to test two of the five sources of BGC variation identified by Lind et al. (2017); namely, whether BGC diversity at the metabolite class level occurs due to gains and losses of entire clusters or, more simply, to shuffling/rearrangement of existing core genes with different tailoring and transport genes. Taking the case of my putative anthraquinone BGCs, I observed that the BGCs were mostly restricted to the *Teloschistales* and a handful of other taxa. However, when I extracted the core PKS gene and searched for homologues within the 45 lichen genomes studied, I found many orthologues and paralogues across all the genomes in the dataset, including taxa that do not produce anthraquinones. It, therefore, appears that the same shared ancestral PKS gene is used for different purposes in different lichen families. Certain paralogous PKS genes show widespread distribution across lichenised fungi (Muggia et al. 2008; Opanowicz et al. 2006), supporting this idea, and orthologous PKSs producing different compounds have been identified in both lichenised and non-lichenised fungi (Singh et al. 2021b; Singh et al. 2022; Kroken et al. 2003; Koczyk et al. 2015). This idea is further supported by the work of Gerasimova et al. (2022), which showed that all non-reducing PKS genes in a dataset of 23 lichen-forming fungal genomes belonged to a single orthogroup. My results, therefore, suggest that for *Teloschistales* anthraquinones, BGC diversification occurred via the rearrangement of existing core PKS genes with different accessory enzymes. To confirm this, I would need to repeat the networking analysis for the additional 301 PKS homologues (represented in Figure 2.3c inset) and assess whether they form part of functional BGCs. If so, I could link them to their metabolite products with the same approach used in this study. There is still another possibility, however, that current orthologue identification approaches may be artificially grouping PKS genes that are structurally similar but not truly orthologous. I believe this to be unlikely, given the stringent filtering used when aligning the orthogroup amino acids and the high support seen in the corresponding gene tree.

Putative anthraquinone BGCs in the *Teloschistales* contain four core genes

Of the 25 *Teloschistales* genomes analysed in this study, 22 contained a putative anthraquinone BGC consisting of four genes: a PKS gene, an EthD gene, a MβL-TE, and an ABC transporter. The *Seirophora villosa* and *Xanthoria* sp. 1 genomes were missing one of the four genes, and the *Caloplaca aegae* genome contained only the core PKS. Whether the additional genes are truly missing or due to fragmented/partially assembled contigs will need to be explored in future work with additional data.

The presence of ABC transporter genes within the BGC is noteworthy. ABC transporters are transmembrane efflux pumps found across the tree of life that actively transport metabolites across cell membranes (Stergiopoulos et al. 2002). In fungi, they are used to expel antibiotics produced by hosts or competitors (Urban et al. 1999) and to pump out endogenously produced metabolites to prevent toxic accumulation (Amnuaykanjanasin et al. 2009). Their ability to transport a broad spectrum of metabolites can also lead to fungal multidrug resistance, a pressing issue in public health (Andrade et al. 2000; Higgins 2007; Fisher et al. 2018).

Anthraquinones are toxic metabolites clearly produced in large quantities by lichens in the order *Teloschistales*. Having ABC transporters adjacent to anthraquinone PKSs could potentially allow for efficient transport of anthraquinones out of the cell before they cause damage to the mycobiont cells and the other symbiont partners. They would also explain how *Teloschistales* lichens can accumulate such large amounts of anthraquinone crystals in the thallus cortex and reproductive structures. As discussed in Chapter 1.1, antitoxin ABC transporters are commonly found within fungal secondary metabolite BGCs (Keller 2015) and have been shown to provide self-protection specifically against the product of the BGC in which they are found (Gardiner et al. 2005). In lichens, ABC transporters have been identified in the usnic acid BGC of *Evernia prunastri* (Pizarro et al. 2020) and two uncharacterised BGCs in *Cladonia uncialis*, one of which was putatively linked to production of the anthraquinone emodin (Bertrand et al. 2018a). The same *C. uncialis* genome was included in this study, but no putative anthraquinone BGCs were identified (Figure 2.3).

Efflux pumps can be costly, diverting energy from growth and DNA repair (El Meouche

et al. 2018). In *Teloschistales* lichens, this may be compensated by their extremely slow growth and the fact that lichen anthraquinones themselves are sun-screening pigments, thereby shielding DNA from UV radiation (Solhaug et al. 1996; Gauslaa et al. 2003). Within the BGC families containing my putative anthraquinone BGCs, only the *Teloschistales* BGCs contained ABC transporters. This may explain why these compounds are so abundant and widespread in this order of lichens. The non-*Teloschistales* genera *Lasallia*, *Umbilicaria*, *Cladonia*, and *Usnea* that are also found in these BGCFs have the core PKS machinery to produce anthraquinones but may be limited in synthesising large quantities due to toxic build-up within the cells and a lack of transporters to expel them. These non-*Teloschistales* taxa may also be synthesising anthraquinones as intermediates in the synthesis of other compounds (as is the case for monodictyphenone synthesis in non-lichenised fungi (Chiang et al. 2010; Neubauer et al. 2016)) and so are converting anthraquinones into other, potentially less toxic compounds before toxic accumulation can occur. Exploring how these ABC transporters are regulated in the *Teloschistales* would allow for a greater understanding of anthraquinone biosynthesis and how it may be affected by environmental or biological factors.

The observed coupling of PKSs with MβL-TEs instead of a TE domain appears to be a common feature of anthraquinone biosynthesis, first described for atrochrysone biosynthesis (Awakawa et al. 2009) and since identified for many other anthraquinones in non-lichenised *Pezizomycotina* (Chiang et al. 2010; Lim et al. 2012; Neubauer et al. 2016; Throckmorton et al. 2016; Griffiths et al. 2016). The phylogenetic analyses of fungal PKSs performed in this study also consistently group these PKSs together along with my putative anthraquinone PKSs (see group V of Li et al. (2010b) and (Liu et al. 2015a) and Figure 2.4b here). Additionally, Bertrand et al. (2018a) observed this structure in the aforementioned emodin BGC from *C. uncialis*. This high sequence-similarity between anthraquinone PKSs of distantly related taxa both at the protein domain level and entire gene cluster level (see Figures 2.4b and 2.3a respectively), suggests that anthraquinone biosynthesis in the *Teloschistales* and other lichens is homologous to that of non-lichenised *Ascomycota* in the *Eurotiomycetes*, *Dothideomycetes*, and *Sordariomycetes* classes. This implies that anthraquinone biosynthesis could be a shared ancestral trait in the subphylum *Pezizomycotina*. A similar hypothesis was proposed for melanin biosynthesis in lichens, where certain lichen PKSs were homologous to melanin PKSs of non-lichen *Pezizomycotina*

and shared a conserved structure of two ACP domains between the PT and TE domains (Pizarro et al. 2020). Ancestral-state reconstruction of anthraquinone type TE-less PKS within the *Pezizomycotina* could explore this hypothesis further and investigate which factors control whether this trait is retained or lost. The fact that homologous anthraquinone PKSs are found between distantly related taxa within subphylum *Pezizomycotina* but not outside the subphylum, gives additional support to the theory that extensive PKS diversification occurred at the root of the *Pezizomycotina* clade via gene duplication and/or horizontal gene transfer (Kroken et al. 2003; Schmitt et al. 2009; Lawrence et al. 2011; Koczyk et al. 2015).

It is worth noting that an alternative strategy for anthraquinone production has been observed in the non-lichenised *Lecanoromycetes* species *Cyanodermella asteris* (Jahn et al. 2017). The BGC responsible for producing the anthraquinone skyrin in *C. asteris* does not contain the TE-less PKS/MβL-TE pairing and shows no apparent synteny with the lichen and non-lichen anthraquinone BGCs discussed in this study. This highlights the importance of linking chemical phenotype to genotype as convergent evolution may lead to different genetic mechanisms producing the same or similar compounds.

Here I focused on the two largest and most diverse families (*Letrouitiaceae* and *Teloschistaceae*) of the four within *Teloschistales*, and the three subfamilies included in the largest family *Teloschistaceae*. The patterns I see are likely to apply to the whole order, however, given the clade's diversity and chemical variability, an essential next step is exploring anthraquinone BGC variation more extensively at the genus level in all families in the order, including the smaller *Brigantiaceae* and *Megalosporaceae* families. Special focus should be placed on *Megalosporaceae* and genera such as *Pyrenodesmia* given that these lineages appear to have secondarily lost anthraquinone pigments. Although these are difficult taxa to sample and sequence at the genome level, including them could allow us to investigate whether secondary pigment loss is due to entire BGC loss, as is the case for usnic acid in the *Parmeliaceae* (Pizarro et al. 2020), individual gene loss (for example loss of ABC transporters), or another genetic mechanism.

2.6 Conclusions

In this study, I implemented a metagenomic approach to assemble and analyse 24 new high-quality *Lecanoromycetes* genomes. My comparative genomic analysis identified high diversity and dissimilarity in lichen biosynthetic gene clusters with most gene clusters being species-specific. Using a similarity-based networking approach I identified a set of putative anthraquinone clusters in the *Teloschistales* with a conserved four-gene structure and high synteny to anthraquinone clusters from non-lichenised fungi. These clusters provide a starting point for future transcriptomic and genetic engineering studies to link gene clusters more definitively to specific anthraquinones and understand how pigment metabolism in lichens is controlled. Based on my analysis, I believe the shuffling of existing PKS genes with key accessory genes such as ABC transporters may explain how *Teloschistales* lichens evolved the capacity to produce such high quantities of anthraquinones without suffering their cytotoxic effects.

2.7 Acknowledgements

I would like to thank Prof. Xavier Llimona for collecting and identifying many of the *Teloschistales* samples. The critically endangered *Teloschistes peruensis* sample was collected with the following permits (Authorization for scientific research of wild flora: R.D.G. No. 441-2018-MINAGRI-SERFOR-DGGSPFFS. Constancy of Deposit to the USM Herbarium: C.D. No. 16-2019-USM-MHN. Export Permit: P.E. N° 003487-SERFOR). I thank Alfonso Orellana, Gwilym Lewis, Justin Moat, Gunnar Ovstebo, Veronica Zuñiga and Miguel Bailetti for collecting the *T. peruensis* sample. I thank Professor Paul Dyer and Dr Igor Grigoriev for granting us permission to use the unpublished JGI Mycocosm genomes of *Xanthoria parietina* and *Usnea florida*. I also thank the Genome Biology and Evolution associate editor Dr Li-Jun Ma and three anonymous reviewers for their valuable feedback on the published paper. I acknowledge the assistance of the HPC Support team at Imperial College London.

2.8 Data availability statement

The genomic data underlying this article are available in the GenBank Nucleotide Database under the BioProject accession PRJNA783777. Individual genome accession numbers are shown in Table 2.1. Genomes downloaded from the public domain are shown in Supplementary Table A.2 along with their relevant databases and accession numbers. Additional data files for anthraquinone PKS amino acid and nucleotide alignments, single-copy orthogroup alignments, raw phylogenetic gene and species trees and HMM seed alignment and tree are deposited in Zenodo doi:10.5281/zenodo.7015195. All bioinformatic scripts used to generate results can be found here: `https://github.com/theo-llewellyn/TeloschistalesMetagenomics`

Chapter 3

Phylogenomics and molecular systematics of the *Teloschistaceae*

3.1 Abstract

The *Teloschistaceae* is one of the most diverse lichen families and displays impressive morphological, chemical and ecological variation. Testing hypotheses about how this trait diversity evolved first requires a robust phylogenetic framework. Previous multilocus datasets have struggled to resolve relationships between genera and subfamilies, which has led to unstable, conflicting taxa across disparate studies and widespread confusion regarding *Teloschistaceae* classification. By combining an 84-taxon genome-scale dataset with densely sampled multilocus data, I could robustly reconstruct previously unresolved intergeneric and subfamilial relationships. This represents the largest genome-scale phylogenetic dataset for a clade of lichenised fungi, both in terms of the number of loci and the number of whole genomes. Phylogenomic analysis showed that around half of the current *Teloschistaceae* genera are not supported and could be combined into existing genera to simplify the current classification. Using a sequential Bayesian dating approach, I could also provide new estimates for *Teloschistales* node ages. I show that genome-scale data dramatically improves the precision of node date estimations and confirm a Late Cretaceous origin of crown *Teloschistaceae* (96-101 Ma), with a peak of genera divergence during the Oligocene. This represents the first fungal study to use whole-genome data for sequential Bayesian time calibration. Therefore, these results provide proof of concept for producing robust time trees from lichen genomes, which can be used in molecular

systematics and future studies on trait evolution.

3.2 Introduction

Lichens display considerable variation in their morphological, chemical, and ecological traits (Honegger 2023). Comparative genomics allows us to understand the genetic basis of these traits. However, to test how trait diversity evolves, we first need a comprehensive phylogenetic tree. As with other fungi, the most common datasets for phylogenetic reconstruction of lichen-forming fungi consist of a combination of the RNA genes ITS, nucLSU, and mitSSU, with one or more protein-coding genes (commonly *RPB1* and *RPB2*). These datasets have informed lichen classifications from subspecies to class level (Miadlikowska et al. 2014).

Although multilocus datasets allow for dense taxon sampling at a relatively low cost, they often struggle to resolve deeper divergences between genera and families. This lack of resolution has led to uncertainty in higher-level relationships in lichen-forming fungi. Genome-scale datasets can help to address this and are just beginning to be applied to lichens (Leavitt et al. 2016; Pizarro et al. 2018; Grewe et al. 2020; Díaz-Escandón et al. 2022). A robust phylogenomic tree can also act as a backbone for multilocus datasets resulting in densely sampled trees with well supported deeper nodes (Varga et al. 2019; Álvarez-Carretero et al. 2022; Chesters 2020).

By resolving deeper nodes, a robust phylogenomic tree is also invaluable for taxonomic work delimiting higher taxa. Taxa are usually delimited through a combination of morphological, biogeographical, biochemical, phylogenetic and ecological traits. Integrating multiple sources of evidence—an approach known as integrative taxonomy—leads to more robust, useable taxa (Dayrat 2005). However, as taxa above species level are generally considered unnatural units, no objective criteria exist for delimiting them beyond monophyly. As a result, their use in comparative studies is limited and higher taxa at the same rank often differ hugely in their trait diversity and species number (Bertrand et al. 2006). This has led to some researchers proposing that rank-based higher taxa should be abandoned entirely (Donoghue et al. 2004; de Queiroz 2006). Despite their lack of uniformity, higher taxon names still allow us to place lower ranks in a wider context and are thus important communication tools.

Given the instability of higher taxa, more objective criteria have been suggested to standardise their delimitation and make them more comparable. Temporal banding, for example, proposes that taxa at the same rank should be a similar age (Avise et al. 1999). However, there is no standard procedure for choosing this temporal band, resulting in a method that can be as subjective as integrative taxonomy without the benefit of delimiting characters for identification. Given the ample discussion on the drawbacks of temporal banding (see Lücking (2019)), I will not investigate this approach here.

Whilst temporal banding and traditional taxonomy assume that species are the only natural ranks, other methods consider all taxonomic ranks within a single evolutionary framework (Humphreys et al. 2014). The Generalised Mixed Yule Coalescent (GMYC) is one such method that has been applied to higher taxon delimitation in mammals, plants, and fungi (Humphreys et al. 2014; Humphreys et al. 2016; Liu et al. 2015b). Promisingly, many GMYC higher taxa, referred to as higher Evolutionary Significant Units (hESUs), match current higher taxa showing that groups delimited by integrative taxonomy already reflect evolutionary reality to some extent. This also suggests that incorporating results from GMYC would not be too disruptive to current classifications. Therefore, whilst branch-length-based methods are still blind to useful delimiting characters, they can provide additional information for clades where the phenotype is not useful.

The widespread use of monophyly as the sole character for delimiting higher taxa, without integrating non-molecular characters, often makes higher taxa impenetrable for field biologists. This is especially apparent in groups with less discerning morphology, such as fungi. In lichens, perhaps the most drastic example is the *Teloschistaceae*, a family with many crustose species generally lacking reliable characters for splitting higher taxa. First described by Zahlbruckner (1898), the family faced almost a century of taxonomic uncertainty (Zahlbruckner 1931; Zahlbruckner 1940) (reviewed by Gaya et al. (2008)), before being settled in an extensive review by Kärnefelt (1989). Kärnefelt systematically provided morphological characters for genera delimitation, and his groupings are still widely accepted and used today.

Although Kärnefelt's delimiting morphological characters were useful, the introduction of molecular phylogenetics suggested that many of them were homoplastic and that groupings were polyphyletic (Gaya et al. 2003; Søchting et al. 2003). The ITS phylogeny of Gaya et al. (2008) confirmed this, especially for the three largest genera *Caloplaca*,

Xanthoria, and *Teloschistes*. A subsequent study by (Gaya et al. 2012) including up to five loci, shed light on higher taxa above the genus level, leading to the proposal of two new subfamilies.

In an attempt to provide higher taxa reflecting the *Teloschistaceae* phylogeny, Arup et al. (2013) proposed a radical new classification, describing 31 new genera and one new subfamily. Many genera lacked clear delimiting morphological characters, which was justified on the basis that extensive phenotypic plasticity had rendered morphology largely uninformative.

Since Arup et al. (2013), over 75 additional *Teloschistaceae* genera and two new subfamilies have been described in multiple studies from different authors (Kondratyuk et al. 2014; Kondratyuk et al. 2015; Kondratyuk et al. 2017; Kondratyuk et al. 2018a; Kondratyuk et al. 2018b; Kondratyuk et al. 2018c; Kondratyuk et al. 2020; Frolov et al. 2020; Bungartz et al. 2020; Wilk et al. 2021; Kondratyuk et al. 2022b; Kondratyuk et al. 2022a; Fayyaz et al. 2022). Most new genera reconfigure existing species rather than accommodate newly discovered ones. This trend of extreme splitting has not been seen in other lichen families, which have much more balanced species-to-genus ratios (Figure 3.1).

Furthermore, re-evaluation of new higher taxa showed some were described from contaminated, misidentified or chimeric sequence data (Llewellyn 2019; Wilk et al. 2021). The tenfold increase in genus names in only 20 years, combined with a growing lack of trust in new *Teloschistaceae* genera, has left most mycologists using the useful but unnatural taxa of Kärnefelt (1989) and classic keys of Clauzade et al. (1985) and Wirth (1987). The *Teloschistaceae* is, therefore, a clear outlier and has been recognised as a priority for lichen taxonomy (Lücking et al. 2017).

In this chapter, I aim to unite the results of disparate molecular studies into a global synthesis of *Teloschistaceae* higher taxa grounded in a robust phylogenomic framework. This chapter will provide the first phylogenomic study of the *Teloschistaceae*, with an aim to resolve relationships between higher taxa and propose recommendations for evolutionary relevant, robust higher taxa that minimise disruption to useful or well-known genera. Given that this will be the most comprehensive *Teloschistaceae* phylogeny to date, I will also use it as an opportunity to re-estimate node ages, as published time-calibrations have only used multilocus datasets with a handful of taxa. Time calibration will allow me to assess whether genome-scale data improves molecular dating estimates compared to

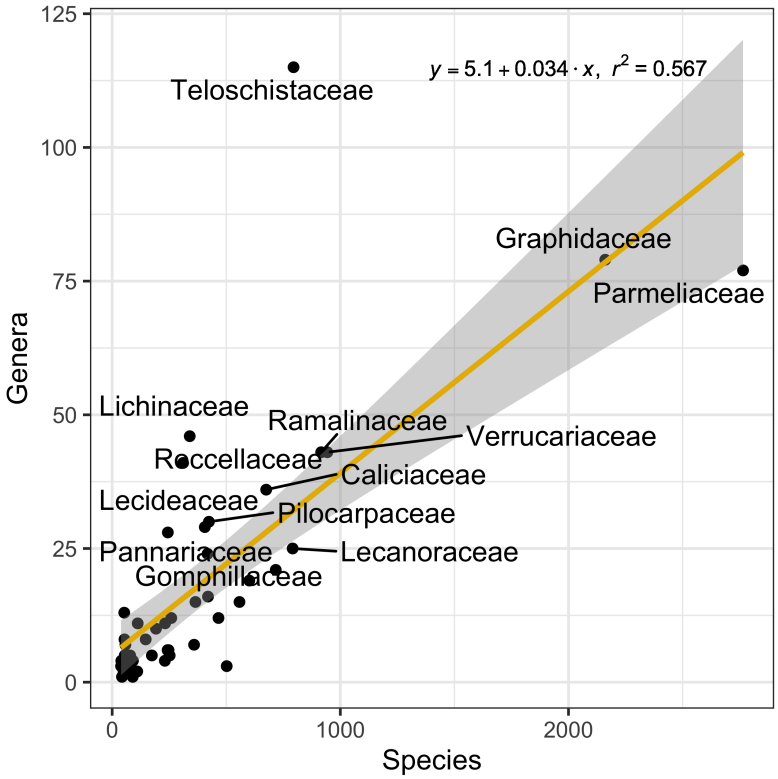


Figure 3.1: Scatterplot showing the number of described species vs genera in the top 50 most diverse lichen families. Data from (Lücking et al. 2017) updated to 2023. Regression line, 95% confidence intervals and equation are shown for linear model: number of genera ~ number of species.

multilocus datasets. As the first lichen study to perform genome-scale molecular dating within a Bayesian framework, rather than with target capture or multilocus data, it will also provide a road map for future studies wishing to do so. Finally, given the lack of reliable phenotypic traits and the difficulty of delimiting higher taxa in the *Teloschistaceae*, I will explore GMYC as an alternative taxon delimitation approach to compare with current integrative taxonomy classifications.

3.3 Materials and Methods

Taxon sampling

I created two complementary datasets for phylogenomic analysis, an exclusively genome-scale dataset of 84 taxa and a multilocus dataset of 552 taxa. For the genome-scale dataset, 57 samples were selected to cover all major lineages within the *Teloschistaceae* family and its sister families *Letrouitiaceae* and *Brigantiaceae*. Additionally, two taxa from the sister order *Caliciales* were sampled to act as outgroups. Supplementary Table

A.3 shows localities and collection information for all genome samples. These samples were added to the 25 *Teloschistales* genomes from Chapter 2 (22 novel, three previously published) to produce the 84-taxon genome-scale dataset.

The 552-taxon multilocus dataset is an updated version of a previous *Teloschistales* multilocus dataset produced by Llewellyn (2019). The previous version of the dataset consisted of 460 taxa covering all four *Teloschistales* families and seven marker loci: the internal transcribed spacer (ITS), mitochondrial small subunit (mtSSU), nuclear large subunit (nucLSU), RNA polymerase II largest subunit (RPB1), RPB2, RET1, and MCM7. The selection of this gene set was based on Gaya et al. (unpublished). To this dataset, I added 92 additional recently published *Teloschistales* sequences, which were downloaded from NCBI. Sample and sequence information for the 460 sequences of Llewellyn (2019) and 92 additional sequences can be found in Supplementary Table A.4.

Genome assembly and annotation

DNA extractions and sequencing for the 59 novel lichen-forming fungal genomes were done as in Chapter 2.3. Lichen metagenomes were assembled and mycobiont contigs isolated as in Chapter 2.3 with one adjustment; after running the first round of BlobTools workflow (Laetsch et al. 2017) I selected and merged *Ascomycota* CONCOCT bins via the BinArena platform (Pavia et al. 2023) rather than in R as it allowed for higher throughput and interactive summary statistics of CONCOCT bin taxonomic assignment. The input file for BinArena was generated by merging GC-content, coverage and taxonomic ID information from BlobTools with CONCOCT bin information using a custom bash script (`https://github.com/theo-llewellyn/TeloschistalesPhylogenomics/blob/main/mycobiont_filtering/BinArena/make_binarena_tsv.sh`). Gene prediction was done as in Chapter 2.3. For subsequent phylogenomic analysis, I filtered genomes for $\geq$ 70% BUSCO completeness to minimise missing data for loci and $\leq$ 5% BUSCO duplicated genes to minimise contamination. Thresholds were chosen based on Bowers et al. (2017)'s definition of high-quality bacterial and archaeal metagenome-assembled genomes (MAGs). Completion was more lenient at $\geq$ 70% rather than >90% as in Bowers et al. (2017) as fungal MAG isolation pipelines are still in early stages of development compared to prokaryote pipelines and are thus less able to isolated fully complete genomes (see Chapter 2.5 for discussion). New whole-genome sequencing data and annotated genome

assemblies are available on GenBank under the BioProject accession PRJNA991097.

84-taxon genome-scale phylogeny

To secure a well-supported phylogenetic backbone, I opted for a sequential approach, first reconstructing a genome-scale phylogenomic tree with the 84-taxon dataset and then using the resulting phylogeny as a backbone constraint for the much more widely sampled 552-taxon multilocus dataset. The 84-taxon dataset allowed me to prioritise dataset width (number of loci) and branch support, whilst the 552-taxon dataset provided dataset length (number of taxa). For the 84 genomes, I identified single-copy gene orthologues using Orthofinder as in Chapter 2.3 and filtered them for genes present in at least ten taxa. I aligned the amino acid sequences using MUSCLE5 (Edgar 2022) and used the corresponding CDS sequences to convert them into nucleotide alignments with PAL2NAL v14 (Suyama et al. 2006). Initial comparisons of trimmed vs untrimmed alignments showed trimming made no difference to concatenated maximum-likelihood phylogenies; therefore, I used untrimmed alignments going forward (See Supplementary Methods A.2 for full details). I split the untrimmed alignment for each gene into two partitions, 1st + 2nd codon positions and 3rd codon position, using the split function of AMAS (Borowiec 2016). I then reconstructed a maximum-likelihood (ML) tree for each gene using RAxML-NG with a GTR+G substitution model (Kozlov et al. 2019).

I used the ML gene trees to filter out genes with outlier long branches. Genes with outlier branches are those where a single branch constitutes over 60% of the total gene tree length. This likely represents a poorly aligned gene or erroneously assigned orthogroup. This produced a final dataset of 1254 loci. I ordered the genes from slow to fast-evolving by calculating the pairwise distance between *Teloschistes chrysophthalmus* (LIQ125TCHR2) and *Xanthoria parietina* (JGI Mycocosm) in each gene tree using the ape function dist.dna in R (Paradis et al. 2019). These two taxa were chosen as they had high-quality assemblies and were represented in all gene trees.

I split the genes evenly across four partitions from slow to fast. Previous work has shown this to be a computationally efficient yet robust method of partitioning hundreds of loci for genome-scale phylogenies (dos Reis et al. 2012; Dos Reis et al. 2014; dos Reis et al. 2015). More partitions may slightly increase precision but would make run times unfeasible. Using only codon positions 1 and 2 for each gene, I reconstructed a concatenated

maximum-likelihood (ML) species tree using RAxML-NG with a GTR+G substitution model and four partitions (Kozlov et al. 2019). Third codon positions were excluded to reduce alignment size to make subsequent Bayesian time calibration feasible. Bootstrapping convergence was tested using the autoMRE criterion within RAxML-NG. I also produced a coalescent-based species tree from the gene trees using ASTRAL-III v5.7.8 (Zhang et al. 2018), with bipartition support assessed via local posterior probabilities (LPP, '-t 3' option in ASTRAL, (Sayyari et al. 2016)). Coalescent methods can account for gene tree discordance due to incomplete lineage sorting, which concatenated approaches can struggle to resolve. Gene-tree concordance was quantified using gene concordance factors (gCF) calculated in IQTree (Minh et al. 2020).

Bioinformatics scripts for gene filtering, ordering and partitioning were adapted from Álvarez-Carretero et al. (2022) for my dataset. Adapted scripts can be found here: `https://github.com/theo-llewellyn/TeloschistalesPhylogenomics/tree/main/raxml84T_tree`.

552-taxon multilocus phylogeny

To use the 84-taxon genome-scale tree as a backbone for the 552-taxon multilocus dataset, I first had to pull the seven marker genes from my whole genome sequences. I used ITSx v1.1.3 (Bengtsson-Palme et al. 2013) to pull full-length ITS sequences from the genomes. For the six other genes, I converted the genome annotations (CDS) into a blast database and then used query sequences from NCBI to search for the genes using blastn via BLAST+ (Camacho et al. 2009). See Supplementary Table A.5 for queries. I filtered the results for matches with an E-value $\leq$0.05 and converted them to fasta files using SeqKit's grep function (Shen et al. 2016). To confirm the extracted sequences belonged to the target mycobiont I blasted them against the NCBI nucleotide database using blastn via the web interface (`https://blast.ncbi.nlm.nih.gov/Blast.cgi`). For ITS, mtSSU and nucLSU, I only kept sequences where the top match was the same species or genus as the target. For protein-coding genes (RPB1, RPB2, MCM7, RET1) I aimed to keep those with top matches to the same family as there are so few published sequences that most of my target taxa are not present in NCBI databases. I also ensured none of the seven marker genes were in the 1254 loci used for the 84-taxon whole genome tree.

For each gene, I concatenated the new sequences pulled from the genomes with the

multilocus dataset sequences (Supplementary Table A.4) and aligned them using MUSCLE5 (Edgar 2022). Here, I chose to trim the seven marker gene alignments before concatenation as the ambiguously aligned regions are likely to have a bigger impact on the concatenated alignment than in the 84-taxon genome-scale dataset. I trimmed alignments using TrimAL v1.4.rev12 with the '-automated1' option (Capella-Gutiérrez et al. 2009). I replaced terminal gaps with '?' and concatenated the seven loci using AMAS (Borowiec 2016). I used modeltest-ng to choose the best substitution model for each gene (Darriba et al. 2020).

Finally, I produced two maximum-likelihood species trees from the trimmed and untrimmed 552-taxon alignments with the 84-taxon genome-scale tree as a constraint. I used IQTree v2.1.2 with an edge-linked partition model and the best substitution model as determined by modeltest-ng. I provided the 84-taxon genome tree with the '-g' option. Branch support was assessed with 1,000 ultrafast bootstrap replicates (UFBoot) (Hoang et al. 2018; Nguyen et al. 2015). I also produced individual gene trees using RAxML-NG, and bootstrapping convergence was tested using the autoMRE criterion. I manually checked gene trees for unusually long branch lengths and reinspected the alignment for potential contaminants or chimeric sequences. These were removed, and alignment, trimming, and multilocus tree inference were repeated.

Time calibration

Akin to the tree inference, I used a sequential-based approach to time calibrate my phylogeny. The approach is based on Álvarez-Carretero et al. (2022) and first estimates divergence times for the 84-taxon genome-scale tree and then uses the posterior node ages as priors to time calibrate the 552-taxon multilocus tree.

I rooted the ML 84-taxon genome tree using *Caliciales* as an outgroup. For node calibrations, I could not use fossils as none exist for this group of fungi. Therefore I used secondary calibrations from Gaya et al. (2015) for the *Teloschistales* (See Figure S6 in their paper). Gaya et al. (2015)'s dated phylogeny is based on eight fossils and two secondary molecular-based calibrations. I also used a secondary calibration from Prieto et al. (2017) for the node connecting *Teloschistales* and *Caliciales*. For node age estimation, I used the Bayesian inference tool MCMCTree v.4.9h/i from the PAML package (Yang 2007). I converted calibrations into 100 Mya units (i.e. 237 Mya becomes 2.37) as MCMCTree can

Table 3.1: Node calibrations used for the time-calibration of the 84-taxon genome-scale tree. 95% confidence interval ages are for crown nodes and shown in 100Mya, i.e. 2.37 = 237 Mya

Node	**95% CI ages**	**Publication**
Caliciales-Teloschistales split	1.37, 2.20	Prieto et al. (2017)
Teloschistales	1.474542, 2.063408	Gaya et al. (2015)
Letrouitineae	0.615328, 1.282252	Gaya et al. (2015)
Teloschistaceae	0.910698, 1.353982	Gaya et al. (2015)
Letrouitiaceae	0.065391, 0.214788	Gaya et al. (2015)
Caloplacoideae	0.6509, 1.12197	Gaya et al. (2015)
Xanthorioideae	0.535822, 0.918282	Gaya et al. (2015)

estimate node ages better in this format. Node calibrations were specified as soft bounds with a default probability of 0.025 that the bounds will be violated. Node calibrations can be found in Table 3.1.

Before running the full dating analysis, I assessed which molecular clock rate model best fit the data using the Bayesian rate model selection tool MCMC3r (Puttick 2019). For this, I initially used the four genes present in all 84 genomes. I generated a list of 32 beta points using MCMC3r to estimate the standard error of the log-marginal likelihood estimate under the strict clock (clk), the independent log-normal (ILN) and generalised Brownian motion (GBM) models. I then ran MCMCTree for each gene separately to estimate node ages using those 32 beta points. Therefore I had 4 genes x 32 beta points x 3 clock models = 384 MCMCTree jobs. I then parsed the MCMCTree output in R and calculated the log marginal likelihood and standard error for each model. I then calculate the Bayes factors and posterior probabilities to identify the best model with MCMC3r.

Initial testing of these four genes showed the GBM model had the highest log-marginal likelihood and posterior probability. Running rate model selection for all 1254 genes would require 120,384 MCMCTree jobs (1254 genes x 32 beta points x 3 clock models). This would require approximately 240 days under maximum parallel conditions on the Imperial High Performace computer (maximum 500 parallel jobs per 24 hours). As this was not feasible within the project timescale, I opted for the GBM model for all 1254 genes. The GBM model generally applies to the tree of life (Tao et al. 2019) and was best fit for other genome-scale time calibration studies (Álvarez-Carretero et al. 2022). It also makes biological sense for my dataset as it assumes descendant branch rates are similar to their ancestral branch rate, especially when the timescale of the branches is short (dos Reis et al. 2016). Given that the *Teloschistaceae* is relatively small (around 1000 species) and all species within the clade are lichen-forming fungi with no lifestyle shifts over short

timescales, I would not expect drastic short-term changes in evolutionary rates which would require the ILN model.

After having determined the best-fit clock model from the four genes mentioned above, I then ran MCMCTree using the full genome-scale dataset (1254 genes). To speed up MCMCTree analysis of the genome-scale dataset, I used the approximate likelihood method (Dos Reis et al. 2011). I then used BASEML with the HKY+G5 substitution model to calculate the branch lengths, gradient and Hessian for each of the four partitions in the 1254 loci concatenated alignment. I then estimated divergence times using the approximate likelihood method and the GBM clock model with a gamma-Dirichlet prior. I specified uniform node age priors by specifying birth-death parameters of 1, 1 and 0.1. For the substitution rate priors, I used a shape parameter (α) of 2 and estimated the scaling parameter (β) using the tree height of the individual gene trees and the divergence time of the root; for this, I calculated the relative branch lengths by dividing largest branch by tree length, which I then used to calculate the average evolutionary rate. Using an α of 2, I calculated β using the equation $\beta = \alpha/\text{average evolutionary rate}$ ($2/0.8840532 = 2.262307$). β calculations were done using an adapted version of 'Estimate_meanrate_beta.R' script, which can be found here: `https://github.com/theo-llewellyn/TeloschistalesPhylogenomics/blob/main/MCMCTree_84T_genome/03_MCMCTree/Estimate_meanrate_beta.R`.

I first ran MCMCTree without the sequence data (i.e. just sampling from the priors) to check that MCMCTree was sampling from the provided priors. I ran MCMCTree five times and combined the outputs into a single trace file. I then ran 16 independent chains with the data using the GBM model. I plotted the results of the 16 chains in R using the MCMCtreeR package v.1.1 (Puttick 2019) to see how the different chains compared. I also compared posterior estimates to the node ages estimated using just the priors. I compiled the 16 posterior chains into a single tracer file and calculated ESS values using the R function effectiveSize from the coda package (Plummer et al. 2006) and Tracer v1.7.2.

I then fitted skew-T distributions to the posterior node age estimates using the R function st.mple from the sn package (Azzalini 2019). I added these distributions onto the tree nodes and reran MCMCTree using the skew-T priors. I did this to check that the skew-T distribution priors overlap with the effective prior MCMCTree uses. I also compared the skew-T priors and effective priors estimated by MCMCTree with the posterior estimates.

This is a key step as it will show if the skew-T distribution priors effectively capture the posterior estimates from the whole genome data. The skew-T distributions will be used as calibration points in the 552-taxon multilocus time calibration.

Once I had ascertained that the skew-T distributions correctly carried the posterior estimates forward, I fitted the distributions for all nodes in the 84-taxon genome-scale tree to the corresponding nodes in the 552-taxon multilocus species tree from section 2.2.4. As with the 84-taxon tree, I used BASEML to estimate the gradient and Hessian for the seven marker loci. I also calculated the substitution rate prior β from seven ML gene trees. I then ran five independent chains without the data, again to verify the prior in MCMCTree. Finally, I ran 32 independent chains with the data using the GBM model and then compared posterior estimates to the node ages estimated using just the priors. I assessed chain convergence as with the 84-taxon tree.

I also calculated node age estimates for the 84-taxon tree with the relative divergence time estimation tool RelTime (Tamura et al. 2012) in MEGA v11 (Tamura et al. 2021). I used both RelTime OLS and RelTime-Branch Length methods. I provided the same secondary calibrations as for MCMCTree with uniform probability and hard bounds, as RelTime does not allow for soft bounds. For OLS, I used the Maximum Composite Likelihood substitution model + G with a gamma parameter of 2, as RelTime OLS does not allow for HKY+G5. For the branch length method, I set the max rate ratio to 100. I then tested the correlation between MCMCTree mean posterior node ages and the two RelTime methods ages using Pearson's correlation coefficient via the cor.test function in R package stats v.3.6.2. I also tested whether node estimates were significantly different using paired two-sided t-tests also in R. Finally I tested whether the size of the confidence intervals between MCMCTree and RelTime were significantly different using paired one-sided t-tests.

I compared my MCMCTree node age estimations for the 552-taxon tree and 95% highest probability density (HPD) confidence intervals with those from Gaya et al. (2015) using a paired two-sided t-test in R for nodes found in both studies to assess the impact of genome-scale data on node age precision.

Higher taxa delimitation

I used the time-calibrated 552-taxon multilocus tree for delimiting higher evolutionary significant units (hESUs) using the Generalised Mixed Yule Coalescent (GMYC) (Fujisawa et al. 2013). hESUs were delimited using the gmyc function of the R package splits (Ezard et al. 2013) with both the single and multiple threshold methods.

3.4 Results

Time calibrated genome-scale phylogeny

To produce a genome-scale time-calibrated phylogeny for the *Teloschistales* order, I sampled and sequenced whole genomes of 59 new taxa using a metagenomic approach. Supplementary Table A.6 shows mycobiont genome assembly statistics. Mycobiont assembly sizes ranged from 22.96 Mbp to 50.51 Mbp (mean 30.60 Mbp). BUSCO completeness ranged from 75.62% to 97.77% (mean 87.92%), and duplicated BUSCOs were all below 1% (mean 0.20%) except *Flavoplaca austrocitrina* (1.52%). Predicted genes numbered from 6,925 to 11,200 (mean 8,076).

I combined these 59 genomes with the 25 *Teloschistales* genomes from Chapter 2 to produce an 84-taxon phylogenomic tree covering 1254 protein-coding genes. Concatenation and coalescent-based species trees were highly congruent, with only the branch separating *Blastenia* A.Massal. and *Gyalolechia* A.Massal. differentiating the two (Supplementary Figure A.6). The ML phylogenomic tree showed 100% bootstrap support for all internodes. All *Teloschistales* families, subfamilies and genera were supported monophyletic groups. Gene concordance factors (gCF) were generally high across the phylogeny (Supplementary Figure A.7). The placement of the *Ikaeria aurantiellina*:*Caloplaca* s.l. sp. 1 clade within the *Xanthorioideae* is a notable exception, with a gCF of 28.3 showing only around 30% of the loci support this relationship. This may also be driving down the gCF of the *Xanthorioideae*:*Teloschistoideae* clade (gCF = 21.8), as both individual subfamilies have high gCFs. Blobplots and BUSCO scores for *Ikaeria aurantiellina* and *Caloplaca* s.l. sp. 1 genomes do not show signs of contamination, suggesting this gene conflict is real and not an artefact of chimeric sequences. I investigated potential long-branch attraction using the SAW method (Siddall et al. 1999), whereby *Ikaeria aurantiellina* was removed

and tree inference repeated to see whether *Caloplaca* s.l. sp. 1 remained in the same position. The same was done by removing *Caloplaca* s.l. sp. 1. Both taxa remained in the same position when the other was removed, suggesting long-branch attraction is not causing their grouping.

I then used the 84-taxon genome-scale tree as a constraint for the 552-taxon multilocus dataset. The 552-taxon phylogeny showed high support along the backbone of the tree. All validly published families and subfamilies are supported. The *Brownlielloideae* and *Ikaeroideae*, were previously shown to be invalid groups of contaminated and misidentified sequences (Llewellyn 2019; Wilk et al. 2021). When these contaminated sequences are removed, the remaining taxa fall into the other three *Teloschistaceae* subfamilies. Therefore, the results presented here confirm that *Brownlielloideae* and *Ikaeroideae* are invalid taxa. The phylogenetic analysis also confidently resolved the relationship between the three subfamilies for the first time, with *Caloplacoideae* supported as sister to *Xanthoriodeae*:*Teloschistoideae* clade. Of the 117 *Teloschistaceae* published genera names represented in the tree, 49 are supported monophyletic groups, 15 are monospecific, 2 are unsupported monophyletic groups, and 51 are reconstructed as either polyphyletic, paraphyletic or nested inside another genus. Therefore around half of the current genera are not supported by our phylogenetic analysis. Figure 3.2 shows the 552-taxon concatenated ML tree with clades collapsed to genus level.

With regard to the dating analyses, for the 84-taxon genome-scale phylogeny, mean node ages estimated by MCMCTree and the two RelTime methods were significantly correlated (RelTime branch length method: Pearson's correlation coefficient $r = 0.98, p < 0.0001$. RelTime OLS method: $r = 0.92, p < 0.0001$). However, whilst mean node ages were correlated between MCMCTree and RelTime results, node age estimates were still significantly different between them, as determined by paired two-sided t-tests (RelTime branch length method: $t_{79} = 10.468, p =< 0.0001$. RelTime OLS method: $t_{79} = 2.5003, p = 0.01448$)(Figure 3.3a and b). The mean difference between MCMCTree and RelTime BL node ages was 7.16 million years. The mean difference between MCMCTree and RelTime OLS node ages was 3.18 million years. 95% confidence intervals were significantly narrower for MCMCTree compared to RelTime, as determined by paired one-tailed t-tests (RelTime branch length method: $t_{79} = 5.5886, p < 0.0001$. RelTime OLS method: $t_{79} = 3.1005, p = 0.001339$) (Figure 3.3c).

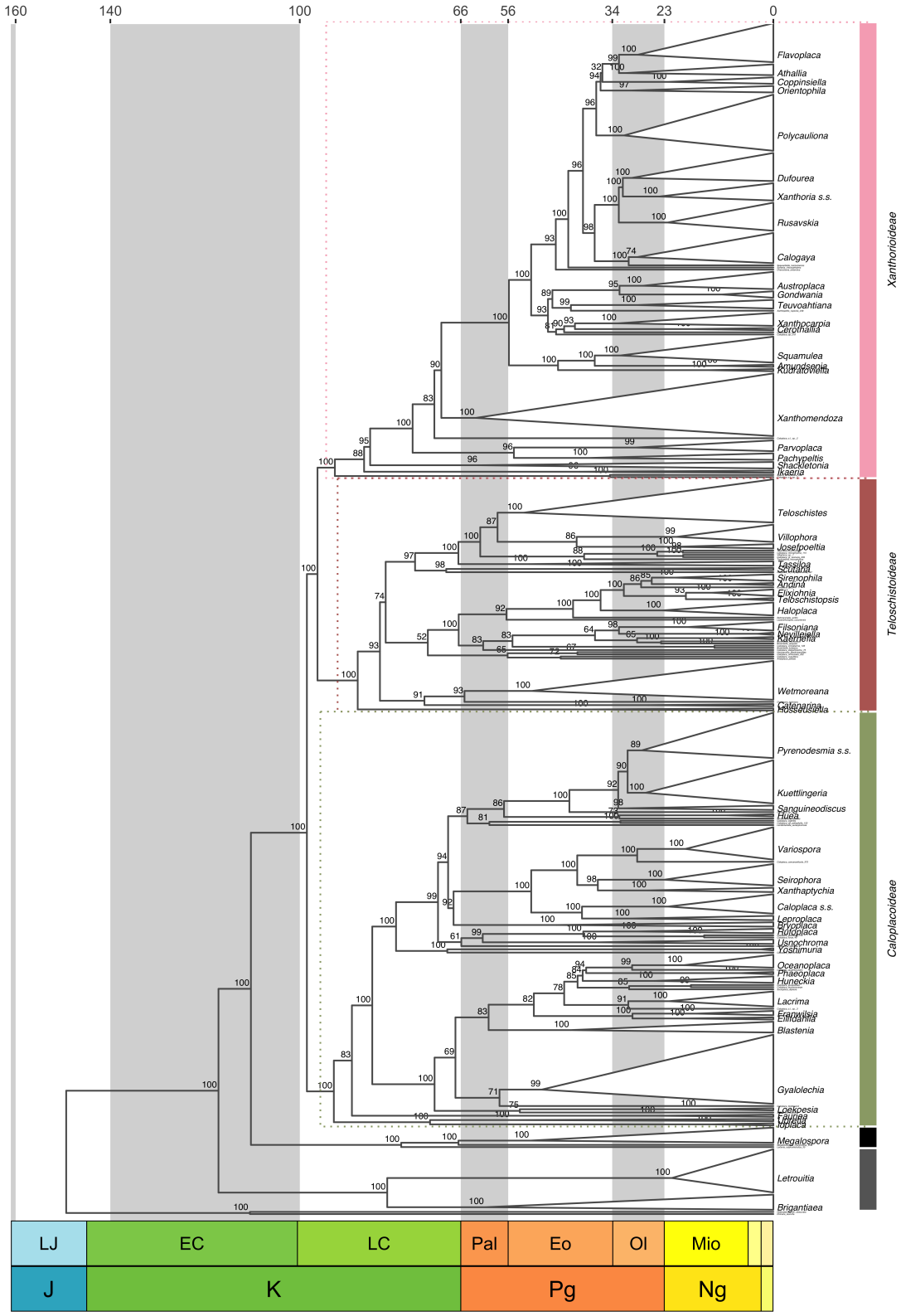


Figure 3.2: Time-calibrated maximum likelihood species tree of 550 *Teloschistales* and two *Caliciales* lichen-forming fungi. Tree inferred from up to 7 concatenated marker loci (ITS, nucLSU, mitSSU, RPB1, RPB2, RET1, MCM7) for 552 taxa with a phylogenomic backbone based on 1254 protein-coding genes for 84 taxa. Nodes are positioned on mean MCMCTree posterior estimates. Ages shown in millions of years before present. Coloured bars highlight *Teloschistaceae* subfamilies. Black bar shows *Megalosporaceae*, grey bar shows *Letrouitineae*. Branches are collapsed for monophyletic genera. Branch labels are bootstrap supports (UFBoot) from IQTree. Genus names indicated in this figure are a combination of old genera (Kärnefelt 1989), the new classification proposed by Arup et al. (2013) and subsequent new genera described in various works.

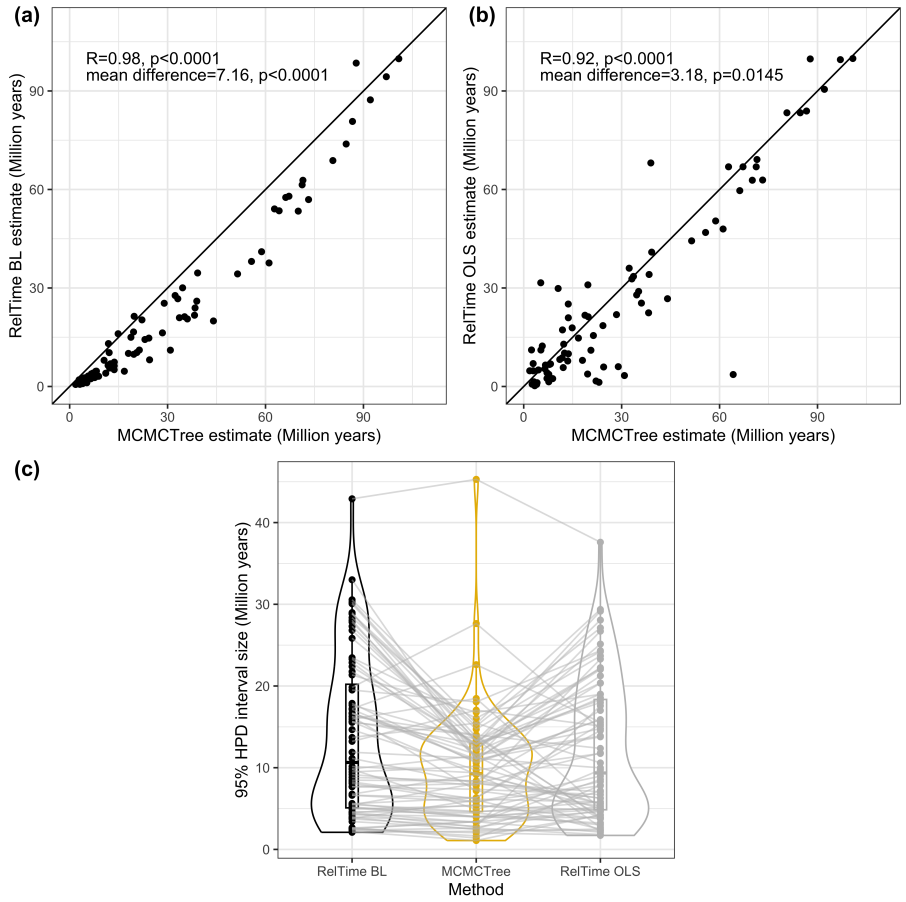


Figure 3.3: Comparison of mean node ages and 95% Highest Posterior Density (HPD) confidence interval sizes between MCMCTree and RelTime estimates using 84-taxon genome-scale data. Scatterplots show a significant correlation between mean estimates from MCMCTree and RelTime branch length method (**a**) and RelTime OLS method (**b**). Pearson's correlation coefficient r and p-values are shown. Paired two-sided t-test mean node age difference and p-values also shown on plots. Diagonal lines are slope=1 intercept=0. (**c**) Violin and boxplots of 95% HPD sizes for node estimates by RelTime branch length method (black), MCMCTree (yellow) and RelTime OLS method (grey). Grey lines connect the same nodes in both methods.

MCMCTree time calibration of the 552-taxon tree estimated that crown *Teloschistales* diverged around 116.2 million years ago (Ma) (95% highest posterior density credibility interval = 107.8-125.5 Ma) during the Early Cretaceous (Figure 3.2). Crown *Teloschistaceae* diverged during the Late Cretaceous around 98.3 Ma (95.6-101.5). Mean node ages for the other *Teloschistales* families *Megalosporaceae*, *Brigantiaceae*, and *Letrouitiaceae* are much younger at 71.7 Ma, 59.2 Ma and 21.6 Ma respectively. These estimates may be younger than the true ages as these other families are species-poor, primarily tropical families and may represent relicts of ancestrally more diverse lineages. Within the *Teloschistaceae*, the three subfamilies diverged at similar times in the Late Cretaceous between 88 and 92 Ma (mean estimates). Supplementary Figure A.8 shows confidence intervals for node ages.

A comparison of the confidence interval sizes in this study with those in Gaya et al.

(2015) showed that 95% HPDs here are generally narrower, and thus node age estimates more precise (Figure 3.4). A paired two-sided t-test for nodes found in both studies showed that 95% HPD interval sizes in this study were significantly different from those in Gaya et al. (2015) at a $p < 0.05$ significance level ($t_{82} = 3.2848, p = 0.001502$). Focusing just on *Teloschistaceae* nodes, I observe the same pattern with significantly narrower 95% HPD intervals for this study compared to Gaya et al. (2015) ($t_{72} = 3.6786, p = 0.0004492$) (Figure 3.4c and d).

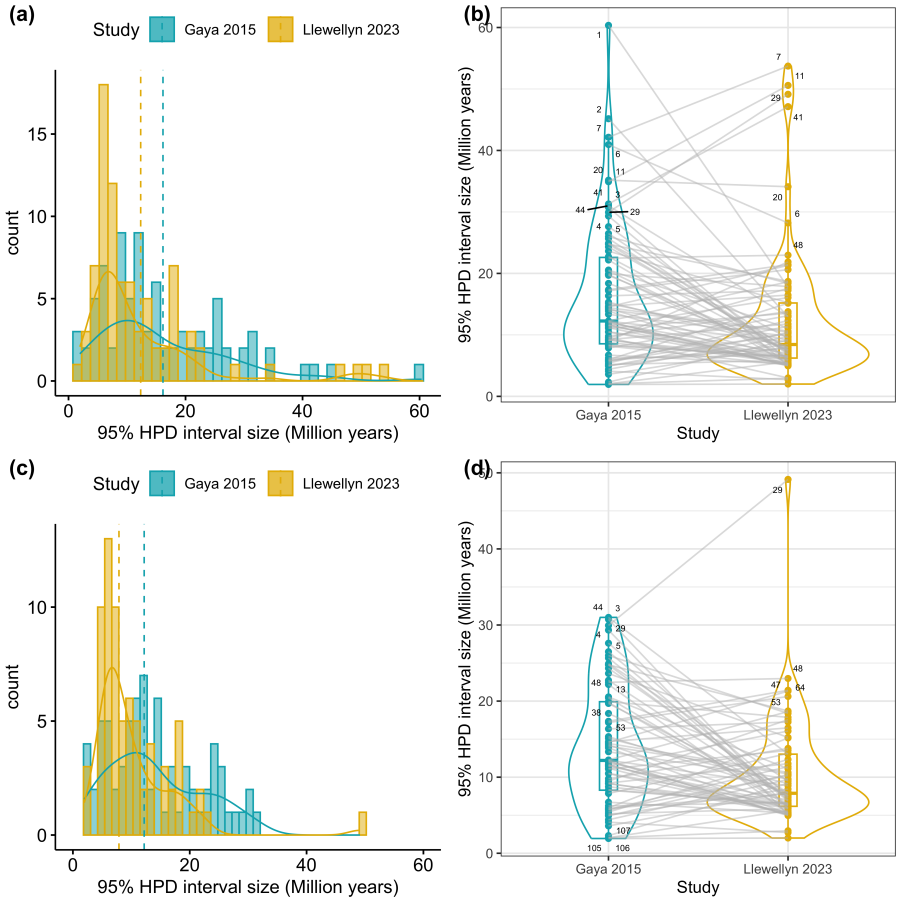


Figure 3.4: Comparison of 95% Highest Posterior Density (HPD) confidence interval size for node ages estimated with multi-locus data in Gaya et al. (2015) (blue) and genome-scale data in this study (yellow). Subplots (**a**) and (**c**) show histograms of 95% HPD sizes for shared nodes in the *Teloschistales* (**a**) and filtered for shared nodes only in *Teloschistaceae* (**c**). Curved line represents the density plot, dashed line shows median. Subplots (**b**) and (**d**) show violin and boxplots of 95% HPD sizes for shared nodes in *Teloschistales* clade (**b**) and *Teloschistaceae* (**d**). Grey lines connect shared nodes in both trees. Outliers are labelled with their node number in the Gaya et al. (2015) tree.

Genera delimitation

GMYC delimited 149 higher evolutionary significant units (hESUs) within the *Teloschistaceae*. Single and multiple threshold methods produced the same results. Figure 3.5 shows how delimited hESUs compare to genera proposed in Figure 3.2 based on monophyly and

correspondence to published genera. GMYC generally delimited hESUs that were smaller than current genera. 30 hESUs match current genera exactly; these are: *Amundsenia* Søchting, Garrido-Benavent, Arup & Frödén, *Aridoplaca* Wilk, Pabijan & Lücking, *Bryoplaca* Søchting, Frödén, & Arup, *Caloplaca* s.s. Th. Fr., *Charcotiana* Søchting, Garrido-Benavent & Arup, *Cinnabaria* Wilk, Pabijan & Lücking, *Coppinsiella* S. Y. Kondr. & L. Lőkös, *Gondwania* Søchting, Frödén, & Arup, *Haloplaca* Arup, Søchting & Frödén, *Harusavskia* S. Y. Kondr., *Huea* C. W. Dodge & G. E. Baker, *Ioplaca* Poelt, *Kaernefia* S.Y.Kondr., Elix, A.Thell & Hur, *Kudratoviella* S. Y. Kondr., L. Lőkös, I. Kärnefelt & A. Thell, *Lacrima* Bungartz, Arup & Søchting, *Lazarenkoiopsis* J.-S. Hur, *Oceanoplaca* Arup, Søchting & Bungartz, *Olegblumia* S.Y.Kondr., L.Lőkös & Hur, *Orientophila* Arup, Søchting & Frödén, *Phaeoplaca* Søchting, Arup, Bungartz, *Rehmaniella* S.Y. Kondr. et al., *Rusavskia* S.Y.Kondr. & Kärnefelt, *Sanguineodiscus* I.V. Frolov & Vondrák, *Seawardiella* S. Y. Kondr., I. Kärnefelt & A. Thell, *Seirophora* Poelt, *Solitaria* Arup, Søchting & Frödén, *Teloschistes* Norman, *Usnochroma* Søchting, Arup & Frödén, *Variospora* Arup, Søchting & Frödén, *Xanthoria* (Fr.) Th.Fr.

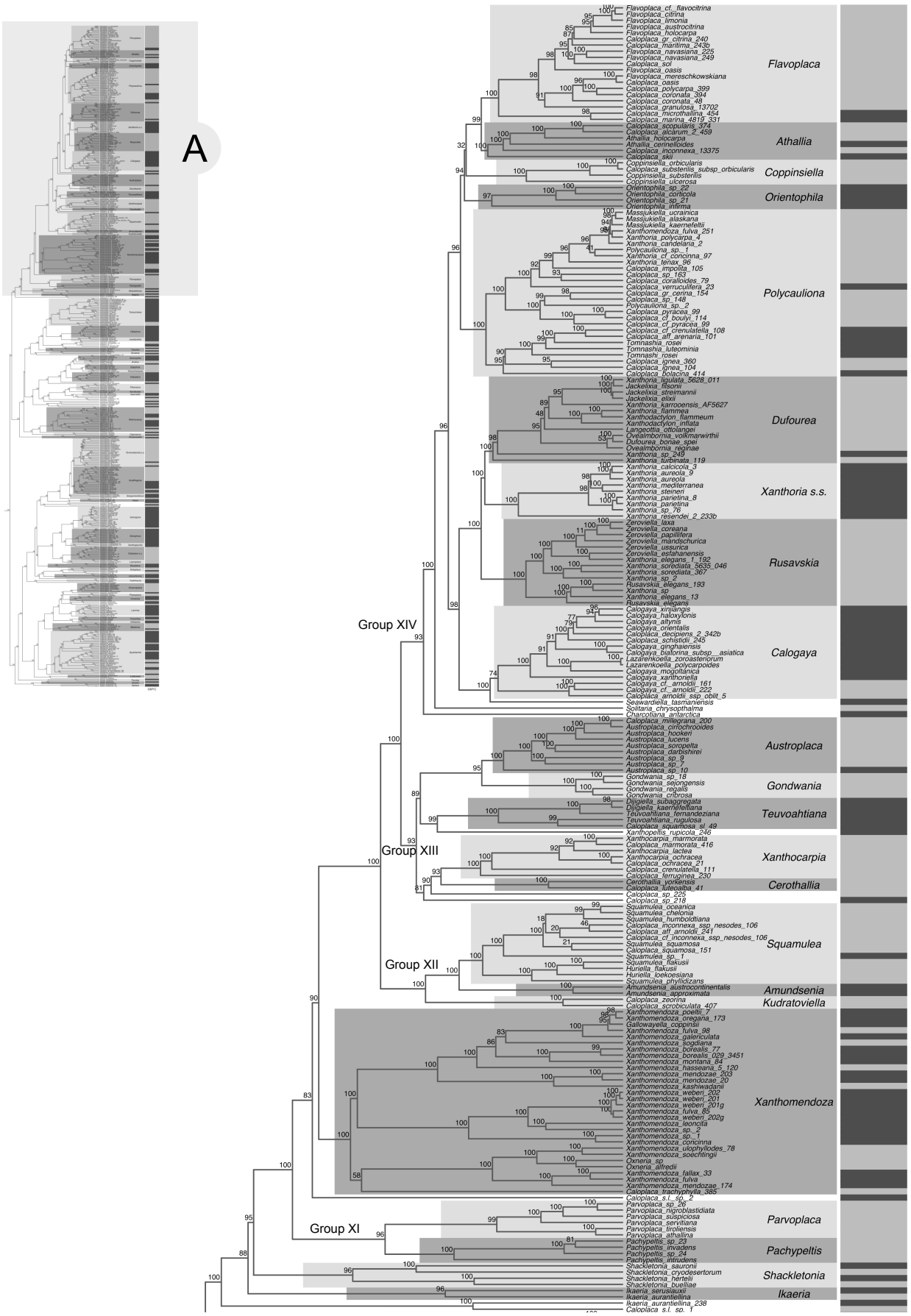


Figure 3.5: (*continued*) subfamily *Xanthorioideae*

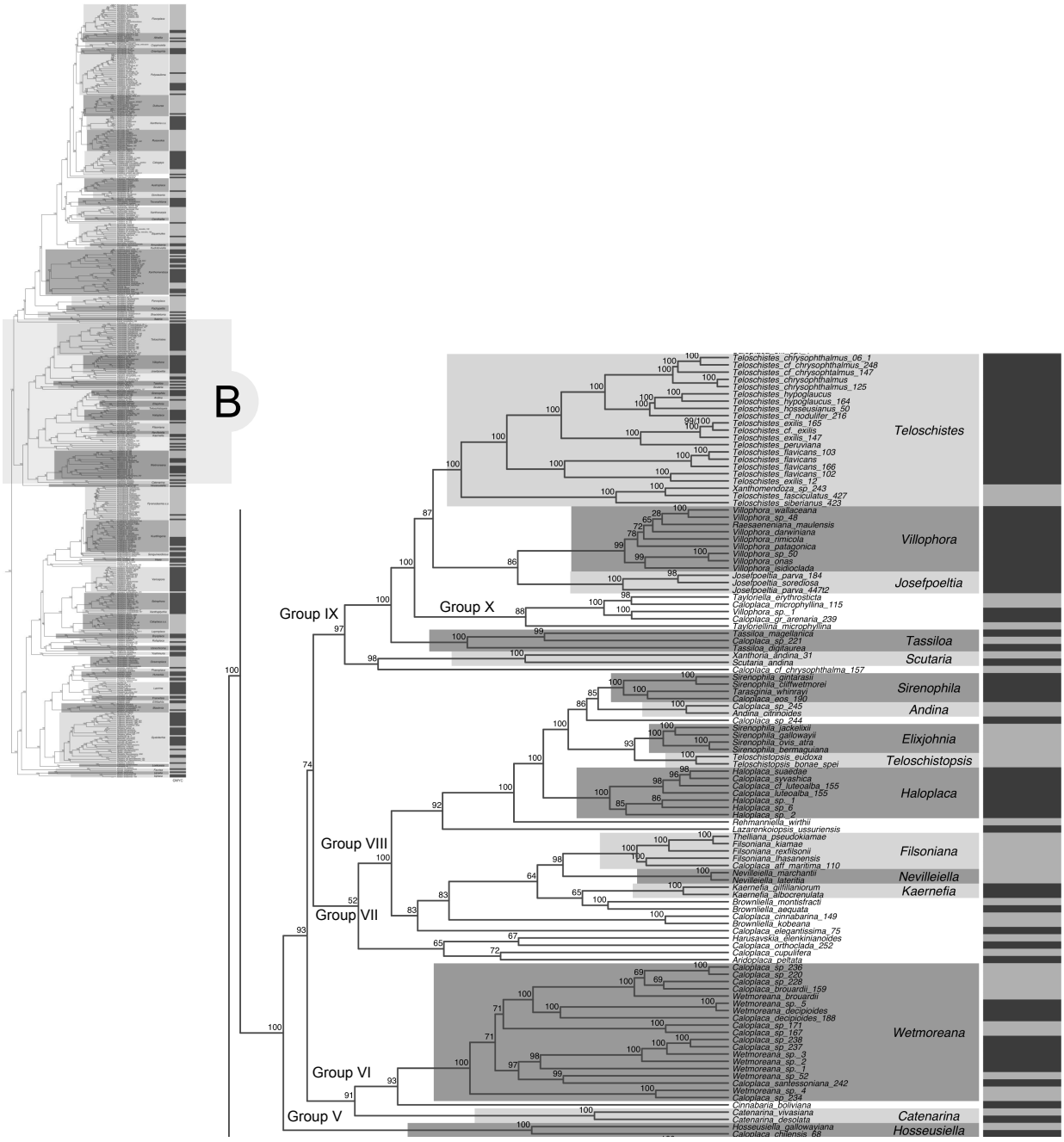


Figure 3.5: (*continued*) subfamily *Teloschistoideae*

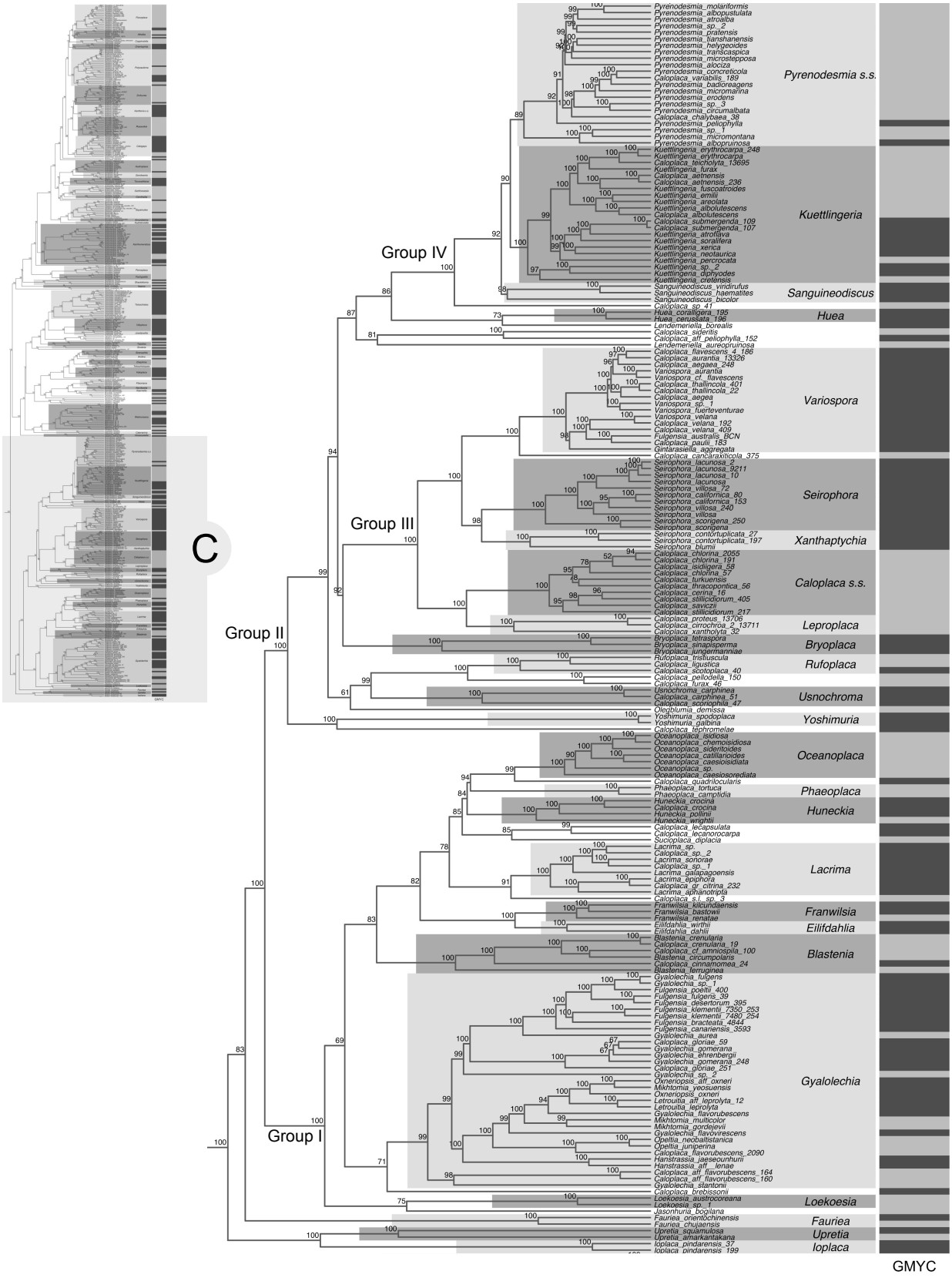


Figure 3.5: Higher taxon delimitation in the *Teloschistaceae*. Supported monophyletic genera are labelled and compared to higher Evolutionarily Significant Units (hESUs) as delimited by GMYC. hESUs and current genera represented by alternating colour blocks. Phylogeny is the 552-taxon time-calibrated phylogeny of Figure A.8 with outgroups removed. Support values are UFBoot bootstraps from IQTree (significant support UFboot >= 95%). Phylogeny split across pages by subfamily for readability.

3.5 Discussion

Here I present the first genome-scale time-calibrated phylogeny for the *Teloschistales* order. My analysis covers 552 taxa in 114 genera, four ingroup families and two outgroup families in the *Caliciaceae*. Using a focused genome-scale dataset as a backbone for a much larger multilocus tree, I robustly reconstructed previously unresolved intergeneric and subfamilial relationships. Time calibration showed that genomic data reduces error ranges of time estimates compared to multi-locus data. Although I did not ask specific time-based questions about *Teloschistaceae* speciation or evolution, the data generated here is a crucial first step for such analyses. Time-calibration also had the purpose of producing more accurate/realistic branch lengths for the subsequent genera delimitation analysis.

Precise node age estimation using Bayesian genome-scale dating

It is now increasingly possible to estimate divergence times using a Bayesian approach with genome-scale datasets. Recent studies have harnessed the power of whole-genomes for Bayesian estimation of divergence times for birds (Jarvis et al. 2014), mammals (dos Reis et al. 2012), insects (Misof et al. 2014), and plants (Harris et al. 2022). For *Fungi*, this has been done for the *Hypocreales* order (Kubicek et al. 2019), the lichen-forming *Lobariaceae* family (Widhelm et al. 2019), and the genera *Aspergillus* and *Penicillium* (Steenwyk et al. 2019). However, given the computational requirements of Bayesian MCMC dating, most of these studies used a limited number of taxa, which can lead to high uncertainty in time estimates.

By applying a sequential dating method (see Álvarez-Carretero et al. (2022) and Harris et al. (2022) for other examples), I was able to estimate node ages with a dataset of 1254 genes and over 500 taxa. This represents the first study to perform sequential Bayesian time-calibration of a fungal phylogeny. To save time and computational resources, I inferred topology and node ages separately, used an efficient partitioning scheme and employed the approximate likelihood method. This way, I could retain the full 1254 gene dataset without having to reduce my dataset to 'clock-like' genes, as is commonly done with genome-scale datasets (Varga et al. 2019; Hill et al. 2022), whilst retaining the benefits of clock model selection and inherent measures of uncertainty for branch lengths and

calibrations (dos Reis et al. 2016). Despite the time-saving solutions applied here, each clock model's analysis still took approximately two months.

Given the time and computational requirements of genome-scale Bayesian time calibration, alternative methods have been developed. These rapid dating methods can be broadly classified into three groups based on their theoretical framework: penalised-likelihood-based methods (Sanderson 2002), least-square dating (LSD) methods (To et al. 2016), and relative rate framework (RRF) methods (Tamura et al. 2012). Of the three, RRF-based tools like RelTime seem to perform best, especially when rates are autocorrelated (Barba-Montoya et al. 2021), and in some circumstances, RelTime produces estimates similar to Bayesian estimation (Mello et al. 2017). Comparing RelTime to MCMCTree results for the 84-taxon dataset in this study showed that although estimates correlated, MCMCTree produced significantly younger estimates with significantly narrower confidence intervals.

In addition to the improved precision of MCMCTree over RelTime, Bayesian inference may be superior for a series of methodological reasons, which I will discuss here. Firstly, Bayesian approaches like MCMCTree do not contain the implicit biases of rapid dating methods regarding how rate variation occurs. Furthermore, rapid methods can inadvertently infer strict clock-like evolution (an unrealistic model in most cases) unless directly specified (see Lozano-Fernandez et al. (2017) and Battistuzzi et al. (2018)). Some of these rapid methods also do not account for uncertainties in branch lengths, fossil calibration densities, or alignment likelihood. The Bayesian approach I use here explicitly factors in these uncertainties. In addition, rapid methods can be run without directly incorporating the sequence alignment, instead reducing this data to branch lengths alone. Therefore, any uncertainty in translating substitutions to branch lengths is lost. Finally, a posteriori RRF calibration methods like RelTime, can struggle to accurately estimate node ages for very short branches (Costa et al. 2022), and under certain complex scenarios, estimate incorrect rate changes and divergence times (Beavan et al. 2020).

A handful of studies on *Fungi* have applied rapid non-Bayesian approaches to genome-scale (Murat et al. 2018; Chang et al. 2019; Miyauchi et al. 2020; Ruiz-Dueñas et al. 2021; Quintanilha-Peixoto et al. 2022; Díaz-Escandón et al. 2022) or target-capture datasets (Widhelm et al. 2021). Of these, Díaz-Escandón et al. (2022) is the only one to feature *Teloschistales* fungi. Their results feature two *Teloschistales* taxa and estimate the

crown node of *Teloschistales* approximately 10 million years earlier than my estimate. Although my estimate fall within their confidence intervals, my results provide much greater precision with a 95% HPD spanning 6 million years compared to their 45 million years. Díaz-Escandón et al. (2022) used the LSD method to calibrate their phylogeny. On top of the general pitfalls of fast dating methods discussed above, the LSD method has specific risks; firstly, it imposes hard-bound fossil calibrations, which can produce artificially precise or even erroneous confidence intervals (Yang et al. 2006). In addition, it does not allow for clock model specification, which hampers confidence interval reliability (Rannala et al. 2007). Finally, LSD confidence intervals are calculated via bootstrap resampling, which is known to produce artificially narrow confidence intervals, as it only quantifies errors related to branch length estimation and not due to rate variation across the phylogeny (Kumar et al. 2016). Despite the longer run times, given the greater precision of MCMCTree demonstrated here and the issues with fast dating methods, I believe a Bayesian approach is still the best option whenever possible.

Given the scarcity and narrow taxonomic focus of genome-scale time-calibrated phylogenies for fungi, I will compare my results to lichen studies that used target-capture or multi-locus data. The node ages estimated here generally overlap with similar-sized groups of lichen fungi in the *Lecanoromycetes*. At 98 Ma, my estimates for the *Teloschistaceae* crown node is comparable to the *Physciaceae* (102 Ma) (Prieto et al. 2017), *Parmeliaceae* (109-112 Ma) (Paz et al. 2011; Divakar et al. 2015), *Icmadophilaceae* (90 Ma) (Prieto et al. 2013), *Lobariaceae* (70 Ma) (Widhelm et al. 2019) and various *Peltigerales* families (100-70 Ma) (Widhelm et al. 2021). Together these point to a common Late Cretaceous origin for many lichen families. Widhelm et al. (2019) linked this to the evolution of epiphytism following the rise of the angiosperms during the Cretaceous Terrestrial Revolution (CTR) (100-70 Ma) (Lidgard et al. 1988; Lidgard et al. 1990). However, we know that diversification in the *Teloschistales* occurred partly due to switching from an epiphytic to a rock-inhabiting lifestyle. Therefore, whilst many lichen families diversified during the Late Cretaceous, they did so for different reasons.

Genus-level node ages overlap with genera in the *Parmeliaceae* (Paz et al. 2011) and *Caliciaceae* (Prieto et al. 2017), with most emerging during the Oligocene and Miocene (see Supplementary Figure A.9), a much cooler period than previous epochs (Zachos et al. 2008). A number of genera also appear to have emerged earlier during the Eocene,

such as *Gyalolechia* (50.3 Ma), *Wetmoreana* Arup, Søchting & Frödén (48.8 Ma), and *Teloschistes* (50.4 Ma). The divergences of *Xanthomendoza* S.Y. Kondr. & Kärnefelt v(64 Ma) and *Shackletonia* Søchting, Frödén & Arup (63.7 Ma) are dated even earlier during the Paleocene. However, these are by no means 'old' in comparison to lichen genera outside *Teloschistaceae* such as *Biatora* and the *Bacidia-Phyllospora* clade, which were estimated to diverge 140–170 Ma during the Jurassic (Beimforde et al. 2014).

I can only make limited direct comparisons of *Teloschistales* node ages as this is only the third study to time-calibrate a *Teloschistales*-focused tree after Gaya et al. (2015) and Freire-Rallo et al. (2023). Freire-Rallo et al. (2023)'s estimates of *Teloschistaceae* node ages focus on *Teloschistaceae* hosts of the parasitic lichenicolous fungus *Tremella caloplacae*. As a result, their dated phylogeny only included 15 *Teloschistaceae* species across six genera. Therefore most nodes in my phylogeny are not present in their tree. For the *Teloschistaceae* crown node, Freire-Rallo et al. (2023) estimate a considerably younger age at 42.2 Ma (25-61.1) compared to my estimate of 98 Ma (95.6-101.5 Ma). This likely reflects the low number of loci they used and the high proportion of missing taxa in their dataset. Nelsen et al. (2020), whilst focusing on the *Ascomycota* as a whole, included seven *Teloschistales* in their study and estimated a very similar crown age for the *Teloschistales* at 113.3 Ma (76.9-151.8) but a considerably younger *Teloschistaceae* age at 60.3 Ma (35.4-92.8).

As my 84-taxon tree used the same secondary calibrations as Gaya et al. (2015), I expected our estimates to overlap. Despite this, my analysis shows that posterior estimates do not simply reflect the prior (Supplementary Figure A.10) thus, the node ages estimated here are informed by my genome and multilocus datasets and can be considered partially independent from Gaya et al. (2015). Gaya et al. (2015) estimated a similar mean node age for the *Teloschistaceae* as here around 98 Ma but with distinctly wider confidence intervals (82.7-113.5 Ma). Therefore my analysis allows for greater certainty of divergence within the Late Cretaceous.

Additionally, statistical tests showed that 95% HPDs in this study are significantly narrower, and thus more precise, than those in Gaya et al. (2015). This difference is most likely due to the use of genome-scale data and wider taxon sampling, although calibration density specification, clock model selection, and software choice may have also had an impact. Regardless of the overall higher precision, I observed four nodes in this study

with markedly wider 95% HPD intervals than other nodes in the tree and also wider HPD sizes than Gaya et al. (2015) (Figure 3.4a and b). Three of these four nodes were in the *Megalosporaceae*, suggesting MCMCTree struggled to obtain precise node estimates for this clade. This may be due to the low number of taxa and lack of genomic data for this clade, meaning that MCMCTree could only use multilocus data to estimate the age of this relatively long branch. Sampling and sequencing *Megalosporaceae* should therefore be a priority going forward.

It is important to note that improved precision does not necessarily imply improved accuracy. Precise dates can still be biased due to methodological biases, incorrect clock model specification and erroneous calibration points. However, if care is taken regarding these parameters, node ages estimated with genome-scale data should theoretically converge on their true values (Rannala et al. 2007; Dos Reis et al. 2014).

Higher taxon delimitation

Phylogenetic analysis supports the monophyly of the three accepted *Teloschistaceae* subfamilies and resolves the relationship between them for the first time (*Caloplacoideae*, (*Teloschistoideae*, *Xanthorioideae*)). As mentioned in the results, gene concordance estimates suggest high gene conflict at the *Teloschistoideae*:*Xanthorioideae* node. This may explain why previous studies have not been able to resolve subfamily relationships confidently.

My phylogeny also resolves many intergeneric relationships. I will discuss these for each subfamily and any major incongruences with current genera delimitation. Any proposed genus circumscriptions are proposed ad interim and are intended as a discussion towards establishing a more evolutionarily relevant classification. They are not to be interpreted as validly accepted nomenclatural changes. This will require extensive taxonomic work, which is outside the scope of this study, and will be the focus of future systematic review.

Caloplacoideae

The subfamily *Caloplacoideae* is the most phenotypically diverse group within the *Teloschistaceae*, ranging from powdery yellow soil-crust lichens in *Gyalolechia* to fruticose grey *Seirophora* species. This clade covers all possible growth habits in the *Teloschistaceae* and presents a high diversity of anthraquinones chemotypes (Arup et al. 2013). It also includes

the type species of the historically largest genus *Caloplaca* and covers broad geographical ranges. Except for a few outliers, species in this clade are generally found at higher latitudes above the equator.

The *Caloplacoideae* contains two large supported lineages (Group I and Group II) and two smaller outgroups (Figure 3.5 Subsection C). The *Upretia*:*Ioplaca* outgroup is sister to the rest of the subfamily on a long branch. Gaya et al. (2012) also identified *Ioplaca* as the first divergence within the *Caloplacoideae*. This is an interesting clade with a distinct Himalayan-centered distribution (Kondratyuk et al. 2018c; Zhang et al. 2019). Similarities between *Ioplaca* and *Upretia* S. Y. Kondr., A. Thell & J. S. Hur were discussed by Kondratyuk et al. (2018c) and are supported here. MCMCTree analysis dates the *Ioplaca*:*Upretia* clade at 71.1 Ma (54.7-86.4) during the Late Cretaceous suggesting this is one of the oldest clades in the whole family. The other smaller outgroup comprises the genus *Fauriea* S.Y.Kondr., L.Lőkös & J.-S.Hur. Although it consistently falls within the *Caloplacoideae* (Kondratyuk et al. 2017; Kondratyuk et al. 2018c; Bungartz et al. 2020), *Fauriea*'s exact position remains uncertain. The length of these branches and the difficulty in placing *Fauriea* suggests a relictual group originating at high altitudes in the Himalayas. Alternatively, the low species number may result from low evolutionary rates and mixing due to their isolated location.

The first large clade within the *Caloplacoideae* (Group I Figure 3.5 Subsection C) consists of the diverse polyphyletic genus *Gyalolechia* sensu Arup et al. (2013), newly described Galapagos genera *Oceanoplaca*, *Phaeoplaca*, *Lacrima* and *Sucioplaca* Bungartz, Søchting & Arup(Bungartz et al. 2020) and a series of smaller genera. Many recently described genera are nested within *Gyalolechia*, some of which are polyphyletic. Following a broad-sense delimitation of *Gyalolechia* by subsuming *Fulgensia* A.Massal. & De Not., *Hanstrassia* S. Y. Kondr., *Opeltia* S. Y. Kondr. et L. Lőkös, *Mikhtomia* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur, *Oxneriopsis* S.Y. Kondr., Upreti & Hur, *Laundonia* S.Y. Kondr., L. Lőkös & Hur and *Elenkiniana* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur, into *Gyalolechia* results in a single supported clade that would reflect the concept of *Gyalolechia* proposed by Arup et al. (2013). However, GMYC analysis delimited 10 hESUs within *Gyalolechia*, suggesting one large monophyletic genus may not accurately represent the branching patterns and evolutionary processes within this group. The Galapagos genera of Bungartz et al. (2020) are all supported by the GMYC analysis suggesting they

are evolutionarily meaningful.

The second large clade within *Caloplacoideae* (Group II Figure 3.5 Subsection C) includes the large genera *Pyrenodesmia sensu lato* A. Massal., *Caloplaca* s.s., *Seirophora* and *Variospora*, a handful of smaller genera and several monotypic branches with species yet to be assigned to new genera. At the base of Group II are *Yoshimuria* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur and *Caloplaca tephromelae*. Although GMYC and phylogenetic analysis support them as a single clade, their lifestyle is very different with *C. tephromelae* being one of the rare examples of a lichenicolous fungus within the *Teloschistaceae*. The next clade to diverge contains *Olegblumia*, *Usnochroma* and *Rufoplaca* Arup, Søchting & Frödén. Although all three genera have full bootstrap support, the whole clade does not.

The next supported divergence within Group II is a clade containing *Leproplaca* (Nyl.) Nyl., *Caloplaca* s.s., *Xanthaptychia* S.Y. Kondr. & Ravera, *Seirophora* and *Variospora* (Group III Figure 3.5 Subsection C). GMYC groups *Leproplaca* and *Caloplaca* s.s. into a single hESU. Since these two genera can be easily distinguished by anthraquinone presence in the thallus, retaining them as separate clades makes sense. GMYC also groups *Xanthaptychia* and *Seirophora* into a single hESU. These two are not readily distinguishable and were separated mainly based on distribution and monophyly (Kondratyuk et al. 2017). Therefore a broader *Seirophora* sensu Arup et al. (2013) may be a more useable delimitation. Within group III, the phylogenetic analysis presented here supports combining *Klauderuiella* S.Y. Kondr. & Hur and *Gintarasiella* S.Y. Kondr. & Hur into *Variospora* sensu Arup et al. (2013). This was also noted by Wilk et al. (2021). This broader circumscription of *Variospora* is also supported by the GMYC analysis.

The divergence following Group III contains *Pyrenodesmia* s.l., *Huea* and a handful of species yet to be assigned to new genera. The order of divergence here is unclear due to a lack of bootstrap support. The genus *Lendemeriella* S.Y. Kondr. is not monophyletic. *Pyrenodesmia* s.l. (Group IV Figure 3.5 Subsection C) is the largest clade within Group III. This was recently split into *Pyrenodesmia* s.s. *Kuettlingeria* Trevis., and *Sanguineodiscus* based on pigment chemistry and molecular data (Frolov et al. 2020). The three genera are recovered here as monophyletic, however, *Pyrenodesmia* s.s. does not have bootstrap support. Although *Pyrenodesmia* s.s. is not supported by bootstraps, given its monophyly, distinguishing morphological characters and unique complete loss of anthraquinones, I believe it should remain sensu Frolov et al. (2020). GMYC splits *Pyrenodesmia* s.l. into

ten hESUs: four for *Pyrenodesmia* s.s., five for *Kuettlingeria* and one for *Sanguineodiscus*. This broadly supports the need to split *Pyrenodesmia* s.l.

The recently described *Caloplacoideae* genus *Fominiella* S.Y. Kondr., Upreti & Hur is of taxonomic concern based on my phylogenetic analysis. The genus contains two species; *F. skii* and *F. tenerifensis* and even when originally described, did not form a monophyletic group (Kondratyuk et al. 2017). *F. skii*, first classified within the *Xanthorioideae* genus *Athallia* Arup, Frödén & Søchting (Vondrák et al. 2012; Arup et al. 2013), was reclassified by Kondratyuk et al. (2017) as outside all three *Teloschistaceae* subfamilies. My analysis places *F. skii* back within *Athallia*. In my results, *F. tenerifensis* originally sat on a long branch within the genus *Seirophora* in the *Caloplacoideae* instead of *Xanthorioideae* with its sister species. Reinspection via BLAST searches of the *F. tenerifensis* nucLSU sequence suggested it may be a misidentified *Seirophora* sample, whilst the mitSSU has top blasts to taxa in the third subfamily *Teloschistoideae*. Without the original sample it is hard to confirm but I suspect *F. tenerifensis* is based on chimeric sequence data. Therefore, I suggest *Fominiella* be considered a nomen invalidum.

Teloschistoideae

The *Teloschistoideae* is dominated by fruticose, foliose and placodiod (lobate) taxa with anthraquinones throughout the thallus. Whilst several cosmopolitan species are included, many subclades show a strong southern hemisphere distribution. *Haloplaca* is a notable exception, showing a coastal distribution in Western Europe.

Phylogenetic analysis shows the *Teloschistoideae* is split into four main clades (Figure 3.5 Subsection B). Three of these are represented in the 84-taxon genome-scale dataset and show over 90% gene concordance, suggesting they are robust, reliable intergeneric clades. The first divergence (Group V Figure 3.5 Subsection B) is the South American endemic *Hosseusiella* S.Y. Kondr., Lőkös, Kärnefelt & A. Thell. Although previous studies could not determine this genus' position within the *Teloschistoideae*, bootstrap support here is high. *Hosseusiella* presents a similar pattern to the *Ioplaca:Upretia* lineage in *Caloplacoideae*, where a small long-branch clade is basal to a more diverse ingroup. However, mean node age for *Hosseusiella* is much younger than the *Ioplaca:Upretia* lineage at 39.1 Ma (11.3-66.6).

Though supported in the 84-taxon tree, the next clade is unsupported in the 552-taxon

tree (Group VI Figure 3.5 Subsection B). Group VI contains *Wetmoreana*, *Cinnabaria* and *Catenarina* Søchting, Søgaard, Arup, Elvebakk & Elix. Figure 3.5 supports previous work suggesting to subsume *Fulgogasparrea* S.Y. Kondr., M.H. Jeong, Kärnefelt, Elix, A. Thell & Hur (*Caloplaca brouardii*, and *C. decipioides*) within a broader well-supported *Wetmoreana* (Arup et al. 2013; Wilk et al. 2021). However, GMYC delimits 11 hESUs within *Wetmoreana*, suggesting narrower genera may more accurately reflect the group's evolutionary history.

The third clade (Group VII Figure 3.5 Subsection B) is also unsupported and consists of one large supported clade (Group VIII) and one smaller unsupported clade. The smaller clade contains a group of monospecific or species-poor genera on long branches. Group VIII contains two large genera *Haloplaca* and *Sirenophila* Søchting, Arup & Frödén and a series of smaller genera. Two of the smaller genera, *Nevilleiella* S.Y. Kondr. & Hur and *Filsoniana* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur, were grouped into a single hESU by GMYC analysis. This suggests a lack of major shift in branching between them; therefore, a broader genus of the two may be more appropriate. *Thelliana* S.Y. Kondr., Kärnefelt, Elix & Hur is nested within *Filsoniana* and thus could be synonymised to give a monophyletic supported *Filsoniana*.

The rest of Group VIII contains *Haloplaca*, *Teloschistopsis* Frödén, Søchting & Arup, *Elixjohnia* S.Y. Kondr. & Hur, *Tarasginia* S.Y. Kondr., Kärnefelt, A. Thell, Elix & Hur, *Andina* Wilk, Pabijan & Lücking, and *Sirenophila*. *Haloplaca* is supported by ML bootstraps and recovered by GMYC. *Teloschistopsis* and *Elixjohnia* are also supported monophyletic groups; however, GMYC groups them into a single hESU. *Teloschistopsis* and *Elixjohnia* are morphologically distinct, with the former having sturdy foliose or fruticose thalli whilst the latter has crustose thalli (Arup et al. 2013). Therefore it makes sense to keep them separate despite the lack of molecular diversity between them. The genus *Sirenophila* is paraphyletic, containing *Tarasginia whinrayi*. *Tarasginia* was originally part of the *Brownlielloideae* subfamily (Kondratyuk et al. 2015), however after excluding contaminant sequence data it repeatedly falls within *Sirenophila* and thus should be synonymised as recommended by Wilk et al. (2021). The newly described Chilean genus *Andina* is sister to *Sirenophila*. In this study, *Andina* forms a supported monophyletic group with an unidentified *Caloplaca* sp. 245, also from Chile. GMYC also groups them into a single hESU. Therefore, I believe *Caloplaca* sp. 245 likely represents an undescribed

Andina species.

The final large clade within the *Teloschistoideae* (Group IX Figure 3.5 Subsection B) contains two large genera, *Teloschistes* and *Villophora* Søchting, Arup & Frödén, and three smaller genera. The type genus *Teloschistes* in the traditional sense is recovered here as monophyletic agreeing with Arup et al. (2013)'s delimitation. This circumscription includes the recently recovered *Niorma* A. Massal. genus, which was recovered to encompass the *T. hypoglaucus* group (Kondratyuk et al. 2013). Given the lack of morphological distinctive characters, I propose here to subsume *Niorma* back into *Teloschistes* sensu Arup et al. (2013). The GMYC analysis also supports this.

Within Group IX (Figure 3.5 Subsection B), *Villophora* is polyphyletic and intermixed with *Raesaeneniana* S.Y. Kondr., Kärnefelt, A. Thell, Elix & Hur, *Josefpoeltia* S.Y. Kondr. & Kärnefelt and *Tayloriella* (nom. illeg. non *Tayloriella* Kylin (Rhodomebaceae, Rhodophyta)). *Raesaeneniana* and *Tayloriella* were originally described as part of the invalid *Brownlielloideae* subfamily (Kondratyuk et al. 2015) and after removing contaminant sequences they fall within *Teloschistoideae*. *Raesaeneniana* was combined into *Villophora* by (Søchting et al. 2021), which is supported here. My results do not support the combination of *Tayloriella* into *Villophora* by Wilk et al. (2021), instead showing a separate monophyletic (albeit not supported) clade containing *T. erythrosticta* and *T. microphyllina* (Group X Figure 3.5 Subsection B) as seen in (Søchting et al. 2021). GMYC suggests 12 hESUs within Group IX, seven of which are monospecific. None of the hESUs in Group IX corresponds to current genera. Sequencing a *Villophora* genome could help to resolve the phylogenetic uncertainty around its monophyly and its placement within Group IX.

Phylogenetic studies repeatedly reconstruct many long-branch monospecific genera within the *Teloschistoideae* (Bungartz et al. 2020; Wilk et al. 2021). GMYC reflects this, delimiting 26 monospecific hESUS here. As a largely southern hemisphere clade, this pattern may be due to under-sampling given the historical bias of lichenology towards Europe and North America. It may, however, represent unequal rates of speciation and extinction within the subfamily. Unequal diversity is pervasive across the tree of life (Rabosky 2010), and given that the phylogeny is now time-calibrated, diversification rate analysis would allow us to test this. Only one study has explored diversification rates in the *Teloschistales*, and they did not detect rate shifts within the *Teloschistoideae* (Gaya

et al. 2015). However, the study focused on the entire order and may not have had the statistical power to detect finer-scale rate shifts within families or subfamilies.

Xanthorioideae

The *Xanthorioideae* is a predominantly Northern Hemisphere clade with some key Southern Hemisphere representatives (Gaya et al. 2008; Arup et al. 2013). It is phenotypically diverse, with crustose, foliose and fruticose growth forms represented. Even within genera, multiple growth forms are seen (e.g. *Xanthomendoza*, *Calogaya* Arup, Frödén & Søchting, *Polycauliona* Hue), suggesting high rates of evolutionary and ecological diversification. The subfamily often dominates littoral habitats, and the microhabitat specificity of certain species has led to them being used to delimit intertidal zones. In contrast to the other subfamilies, *Xanthorioideae* has notably uniform secondary metabolite profiles, with almost all species displaying one of two chemosyndromes (Arup et al. 2013).

From a phylogenetic perspective, the *Xanthorioideae* subfamily is relatively imbalanced, with a stair-like topology (Figure 3.5 Subsection A). Although the subfamily as a whole is supported, some basal long-branch divergences are not. Thus the eight basal GMYC hESUs delimited that cover the genera *Ikaeria* S.Y. Kondr., Upreti & Hur and *Shackletonia* are unreliable and further work will be needed to resolve them.

Following the divergence of *Ikaeria* and *Shackletonia* is a supported clade containing the crustose genera *Pachypeltis* Søchting, Arup & Frödén and *Parvoplaca* Arup, Søchting & Frödén (Group XI Figure 3.5 Subsection A). Group XI consistently emerges in phylogenetic analysis; however, its placement within *Xanthorioideae* is unclear. Previous studies resolved it as sister to the foliose genus *Xanthomendoza* (Arup et al. 2013; Wilk et al. 2021), whilst Figure 3.5 shows it diverging prior. GMYC recovers Group XI as a single hESU. The original description of the two genera suggested they could be merged (Arup et al. 2013). However, given their individual bootstrap support and distinct morphology and ecology, I believe keeping them separate reduces unnecessary taxonomic changes.

The next clade to diverge is the species-rich *Xanthomendoza* genus. My results support a broad circumscription of *Xanthomendoza* sensu Arup et al. (2013). The newer genera *Oxneria* S.Y. Kondr. & Kärnefelt, *Gallowayella* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt, Elix & A. Thell, and *Golubkovia* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur Kondratyuk et al. (2014) are all nested within *Xanthomendoza*. As *Xanthomendoza*

is the earlier genus name, the other genera should be subsumed into *Xanthomendoza*. This has the additional benefit of reducing taxonomic complexity. However, GMYC splits *Xanthomendoza* into 12 hESUs suggesting subclades may be under different evolutionary pressures. The placement of *Xanthomendoza* is also uncertain. Although studies consistently resolve it as a basal lineage within the subfamily, gene conflict is high at this node (Figure 3.5 Subsection A). Gene conflict can result from interesting evolutionary processes such as hybridisation and horizontal gene transfer (Degnan et al. 2009). Whilst research into these processes in lichen-forming fungi is still in its infancy, genome-scale data is already revealing evidence of hybridisation in similar-sized genera from other lichen families (Keuler et al. 2020; Ament-Velásquez et al. 2021; Keuler et al. 2022). As the oldest genus in the family (64 Ma) and a dominant lichen group in many Northern Hemisphere systems, further investigation into the evolutionary history of *Xanthomendoza* is a potentially fruitful area of future research.

After the *Xanthomendoza* divergence, the rest of the backbone is well supported. The *Kudratoviella*:*Amundsenia*:*Squamulea* clade diverges first (Group XII Figure 3.5 Subsection A). *Huriella* S.Y. Kondr. was recently synonymised with *Squamulea* Arup, Søchting & Frödén by Bungartz et al. (2020), due to a lack of differentiating morphology. GMYC, however, separates them into different hESUs, suggesting they are distinct phylogenetic clusters. This provides another example where a useful classification conflicts with an evolutionarily relevant one.

Following Group XII, the *Xanthorioideae* diverges into two large supported clades containing multiple genera. The first clade (Group XIII Figure 3.5 Subsection A) contains *Cerothallia* Arup, Frödén & Søchting, *Xanthocarpia* A. Massal. & De Not., *Xanthopeltis* R. Sant., *Teuvoahtiana* S.Y. Kondr. & Hur, *Gondwania* Søchting, Frödén & Arup and *Austroplaca* Søchting, Frödén & Arup. Within Group XIII, *Dijigiella* S.Y. Kondr. & Lőkös is nested in *Teuvoahtiana* and could thus be synonymised. This is supported by GMYC, which also includes the monospecific *Xanthopeltis* in the same hESU. *Xanthopeltis* is a South American endemic with unusual peltate thallus morphology. Due to its distinctness, I believe it should remain a monospecific genus.

The second large clade (Group XIV Figure 3.5 Subsection A) contains three monospecific genera *Charcotiana*, *Solitaria*, and *Seawardiella*, which are supported by GMYC, and nine larger genera. Within Group XIV, *Lazarenkoella* S.Y. Kondr., Kärnefelt, A. Thell,

Elix & Hur is nested within the species-rich *Calogaya* genus. Given that *Calogaya* is morphologically well defined, subsuming *Lazarenkoella* into *Calogaya* is preferable. *Zeroviella* S.Y. Kondr. & Hur is another nested genus in Group XIV that should be synonymised with *Rusavskia*. GMYC also suggests a single hESU for this group. The genus *Martinjahnsia* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt, Elix, Hur & A. Thell was described to incorporate *Xanthoria resendei*, which sits as a sister species *Xanthoria* s.s. Although *Xanthoria* s.s. and *Martinjahnsia* are both monophyletic, as *Xanthoria* s.s. is well-defined and one of the most widely used and recognised genus names of the family, I believe combining *Martinjahnsia* into *Xanthoria* would simplify the classification and facilitate communication. GMYC also supports this grouping. The topology of this clade in Figure 3.5 also supports Arup et al. (2013)'s synonymisation of *Ovealmbornia* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt, Elix & A. Thell, *Xanthokarrooa* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt, Elix & A. Thell, *Jackelixia* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt & A. Thell, and *Xanthodactylon* P.A. Duvign. with *Dufourea* Ach. Including new sequence data here shows *Langeottia* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur can also be subsumed into *Dufourea*.

Finally, within Group XIV, the genus *Polycauliona* is shown to contain the genera *Tomnashia* S.Y. Kondr. & Hur, *Igneoplaca* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur, *Verrucoplaca* S.Y. Kondr., Kärnefelt, Elix, A. Thell & Hur and *Massjukiella* S.Y. Kondr., Fedorenko, S. Stenroos, Kärnefelt, Elix, Hur & A. Thell. As these names postdate *Polycauliona* and describe taxonomic groups that are either unnatural or not meaningful, they should be subsumed into a broader *Polycauliona* sensu Arup et al. (2013). *Polycauliona* is a species-rich clade mostly found in southwestern North America. It is also one of the most morphologically diverse genera in the family, showing crustose, lobate, subfruticose and foliose thallus growth forms. Given its high species number and phenotypic diversity, it is surprising that *Polycauliona* is estimated to have diversified only 31.1 Ma (28.4-33.7) (Figure 3.2). Diversification rate analysis could explore whether this represents a burst in speciation that could be linked to adaptive radiation.

My time-calibrated phylogeny opens the door for future biogeographical analyses to test whether the divergences of observed clades could be related to or explained by geographical events. For example, with *Teloschistoideae* genera occurring in Southern Africa, South America, Australia and Antarctica, it is tempting to suggest a Gondwanan origin

of this subfamily. However, with a mean node age around 80 Ma, the divergence time of this group significantly postdates the splitting of the Southern Hemisphere landmasses (165-150 Ma) (Sanmartíin et al. 2004), suggesting dispersal rather than vicariance explains their current distribution. Biogeographical analysis may also shed light on the observed long-branch monospecific genera; if the observed pattern is true and not simply due to undersampling, I would hypothesise that monospecific taxa are more likely to be range-restricted endemics or to occupy relictual habitats. Under this hypothesised scenario, previously diverse lineages occupying large areas underwent a process of gradual extinction as their habitat was lost, leaving relicts in increasingly isolated patches. This pattern was observed in pygmy chameleons where long-branch lineages with few sister taxa occupy fragmented forest patches within an increasingly arid South Africa (Tolley et al. 2008).

Currently, accurate distribution data are lacking for most species within the *Teloschistaceae*, especially in the Southern Hemisphere. Therefore, an emphasis on recording, including for common species, is essential to understand the evolutionary history of this family. Further sampling and better estimates of missing data for each lineage would also allow for more accurate diversification rate analyses. The genus *Teloschistes* presents a promising case study for exploring biogeographical questions as it contains both widespread, cosmopolitan species such as *Teloschistes chrysophthalmus* and hyper-range restricted species such as *Teloschistes peruensis*. Furthermore, *Teloschistes* species are more conspicuous and generally easier to identify and record than crustose *Caloplaca* s.l.

3.6 Conclusions

Here, I have shown that genome-scale data of lichen-forming fungi dramatically improves the precision of node date estimations compared to multi-locus estimates. I confirm with high precision a divergence time of crown *Teloschistaceae* during the Late Cretaceous between 96 and 101 Ma, with a peak of genera divergence during the Oligocene. Phylogenomic analysis showed that around half of the current *Teloschistaceae* genera are not supported and could be combined into existing genera to simplify the current classification. However, since the Late Cretaceous, the group has undergone extensive genetic and phenotypic diversification, and the classification should reflect this. I found evidence for 149 supported hESUs in the family, with 31 matching current genera. These 31 genera

can, therefore, be considered robust higher taxa reflecting both traditional integrative taxonomy and evolutionary processes. The remaining genera must be settled case-by-case, considering a trade-off between usability and biological reality. Through this approach, we can work towards a *Teloschistaceae* classification that promotes communication and benefits users whilst still being grounded in evolutionary history.

Chapter 4

Paired genome-metabolome analysis reveals a complex interplay between secondary metabolite profiles, biosynthetic gene cluster variation and phylogenetic relatedness

4.1 Abstract

Over 1000 secondary metabolites have been identified from lichen-forming fungi, many of which are found nowhere else in nature. These secondary metabolites are encoded by tightly linked genes called biosynthetic gene clusters (BGCs), which are highly variable and show extensive turnover between even closely related taxa. Analyzing paired genome-metabolome datasets allows us to link biosynthetic gene clusters with their respective metabolites and has greatly enhanced natural product discovery. However, very few studies consider this type of analysis within a phylogenetic context to understand how these metabolites evolved. Here I examine genomic and metabolic variation within an evolutionary trait-based framework using new metagenomes and untargeted metabolomes

for 79 lichen-forming fungi in the *Lecanoromycetes* class. I found that when summarising metabolome diversity into dissimilarity indices between species, secondary metabolite variation did not correlate with biosynthetic gene cluster variation or phylogenetic relatedness. However, presence-absence patterns of individual metabolites show phylogenetic signal and closely related taxa tend to have similar-sized metabolomes. I also specifically investigated the evolution of anthraquinones, an abundant metabolite group in the *Teloschistales*. Ancestral state reconstruction of anthraquinone BGCs suggests a scenario of BGC evolution in this group of lichens consisting of evolutionarily stable core genes and dynamic lineage-specific accessory genes. Joint modelling of anthraquinone BGCs and metabolites suggest changes at the BGC level do not appear to have increased anthraquinone chemical diversity within the *Teloschistales* and may instead be linked to regulation and transport. My results highlight the importance of paired genome-metabolome datasets for understanding metabolome evolution and suggest a complex interplay between genome and metabolome, rejecting the idea of a one-to-one relationship between BGCs and lichen metabolites.

4.2 Introduction

Lichens produce an impressive array of secondary metabolites, with over 1000 compounds catalogued (Molnár et al. 2010). Only around 5% of lichen compounds are found in other organisms (Elix et al. 2008), which suggests lichens have undergone significant metabolic diversification and specialisation. The diversity of lichen compounds, together with the fact that they can constitute up to one-fifth of a lichen's dry weight (Shrestha et al. 2013), suggests they have been key for adapting to new ecological niches and changing environments. Their photoprotective properties, in particular, have also contributed to the success of lichens in extreme, exposed ecosystems that many other organisms find inhospitable (Gaya et al. 2015).

Lichen metabolites are often taxon-specific, which has led to their widespread use in classification and identification. Chemotaxonomy is especially useful for lichen-forming fungi as many clades provide few morphological distinguishing characters. However, beyond species level, there is considerable variation in the phylogenetic signal of secondary metabolites, with some studies reporting strong correlations between shared metabolic content and phylogenetic relatedness (Miadlikowska et al. 2000; Tehler et al. 2001; Schmitt

et al. 2004; Lücking et al. 2008; Crespo et al. 2010), whilst others do not (LaGreca 1999; Blanco et al. 2006; Nelsen et al. 2008; Lutsak et al. 2017). Assessing the phylogenetic significance of lichen secondary metabolites is limited by the accuracy and sensitivity of widely used, traditional metabolite identification methods like thin layer chromatography (TLC). Tools such as liquid chromatography-mass spectrometry (LCMS) are less error-prone and can detect low abundance metabolites and highlight subtle chemical differences between taxa (Olivier-Jimenez et al. 2019). For example, the widely accepted notion that *Cladonia rangiferina* lichens cannot produce usnic acid was recently shown to be an artefact of the low sensitivity of TLC (Egbert et al. 2022).

As discussed in Chapter 2, secondary metabolite biosynthetic pathways in fungi are encoded by sets of adjacent, tightly linked genes known as biosynthetic gene clusters (BGCs). Lichen mycobiont genomes from the *Lecanoromycetes* class (*Ascomycota*) contain high BGC numbers and are particularly enriched in type I PKSs (Gill et al. (2023) and references within). BGCs also often outnumber the known compounds for a species, suggesting a huge metabolic diversity waiting to be described. My results from Chapter 2 support a growing body of research, showing that BGCs are hotspots of genome evolution in fungi with high turnover rates even between closely related species (Lind et al. 2017; Vesth et al. 2018; Gluck-Thaler et al. 2020; Robey et al. 2021; Singh et al. 2022). However, we know very little about how these changes at the genomic level translate into secondary metabolite variation. Establishing this link would allow us to understand which evolutionary processes shape lichen secondary metabolomes.

One approach for comparing genomes and metabolomes is multi-omics, i.e. generating both genomic and metabolomic data for a given set of samples. Over the past decade, paired genome-metabolome datasets have become increasingly common for bacteria (Duncan et al. 2015; Goering et al. 2016; Maansson et al. 2016; McClure et al. 2016; Männle et al. 2020; Soldatou et al. 2021) and are beginning to emerge for fungi too (Schmidt et al. 2018; Hicks et al. 2021; Gonçalves et al. 2022; Zhang et al. 2022; Duan et al. 2023; Hebra et al. 2022; Caesar et al. 2023). Whilst paired genome-metabolome studies greatly enhance natural product discovery for bioprospecting, very few consider the results within an evolutionary framework. To understand how metabolomes and their associated BGCs evolve, a comparative approach is required where data from a group of related species are compared within their phylogenetic context. For fungi, a handful of comparative studies

exist at both inter and intraspecific levels. In these studies, patterns of variation in BGC repertoires were reflected in secondary metabolite variation (Higa et al. 2020; Witte et al. 2021; Witte et al. 2022), suggesting a direct link between genomic changes and metabolite profiles. In clades with highly divergent metabolite profiles, the frequent gains and losses of metabolites have been linked to BGCs associating with dynamic parts of the genome such as accessory chromosomes and subtelomeric regions (Witte et al. 2022; Witte et al. 2021).

For this chapter, I generated paired genome-metabolome data for 79 lichen taxa within the *Lecanoromycetes*, which I analysed within a phylogenetic framework to explore the relationship between genomic variation, secondary metabolome variation and phylogenetic relatedness. Specifically, I aim to test whether patterns of BGC variation correlate with patterns of metabolite variation and if that variation can be explained by phylogenetic relatedness. Since BGC turnover is so high in this clade, I expect metabolites to show a similar level of variation. As well as BGC/metabolite identity—in other words which metabolites are produced—I aim to test whether the overall number of BGCs and metabolites per species correlates with phylogeny. Finally, I aim to explore whether patterns at the global metabolome level are reflected in specific metabolites and their respective BGCs. For this, I focus on anthraquinones within the *Teloschistales* as a case study, given that they are the dominant metabolites for the group, and the putative BGC is now characterised. For this last section, I aim to elucidate how anthraquinone BGCs evolved and test whether understanding their evolutionary history can shed light on anthraquinone variation, both within the *Teloschistales* and between *Teloschistales* and its sister clades.

4.3 Materials and Methods

Genome sequencing, assembly and annotation of eight additional taxa

Samples for whole genome data are those presented in Chapters 2 and 3, with an additional eight non-*Teloschistales* taxa covering the families *Arthoniaceae* (order *Arthoniales*), *Parmeliaceae* (*Lecanorales*), *Psoraceae* (*Lecanorales*) and *Pyrenulaceae* (*Pyrenulales*). These non-*Teloschistales* samples were selected to cover other lichen clades reported to produce anthraquinones. I sequenced and assembled the novel taxa using short-read Illumina sequencing as in Chapters 2.3 and 3.3. I predicted mycobiont genes and identified

orthologues as in Chapter 2.3. I annotated secondary metabolite gene clusters (BGCs) using antiSMASH (Blin et al. 2019) and grouped them into BGC families using BiG-SCAPE (Navarro-Muñoz et al. 2020) as in Chapter 2.3. Supplementary Table A.7 shows sample localities and collection information, and Supplementary Table A.8 shows assembly quality metrics. New whole-genome sequencing data and annotated genome assemblies are available on GenBank under the BioProject accession PRJNA991097.

Phylogenomic tree reconstruction for the full dataset was done as in Chapter 2.3, i.e. filtering for 50% single-copy orthologues, aligning with MAFFT, trimming with trimAl and concatenated tree reconstruction with IQTree.

Metabolite extraction and LCMS/MS analysis

The Tandem Liquid Chromatography Mass Spectrometry (LCMS/MS) analysis described in this section was performed in collaboration with Dr Pierre Le Pogam Alluard, Department UMR CNRS 8076 BioCIS at Université Paris-Saclay, France.

For each sample that had been genome sequenced, I also performed secondary metabolite extractions. Samples were from collections, stored at room temperature in the dark. All samples were under five years old except three that were between five and ten years old. I transferred approximately 10 mg of clean lichen thallus to 2 ml glass LCMS vials and ground them using a sterile plastic pestle. I added an appropriate volume of LCMS grade acetone to achieve 10 mg sample/0.5 ml acetone. I vortexed vials thoroughly every 5 minutes for 30 minutes. I centrifuged tubes for 2 minutes at 10,000 RPM and transferred the liquid extract to clean 2 ml glass vials. Acetone was evaporated within a fume hood overnight for shipping and redissolved upon arrival at Department UMR CNRS 8076 BioCIS, Université Paris-Saclay.

We analysed metabolite extractions using an Agilent 6546 Accurate-Mass Q-TOF hyphenated with a 1290 Agilent Infinity II LC system. The chromatographic system was fitted with a Zorbax RRHD Eclipse Plus C18 column (2.1x50mm, 1.8 μm). Elution solvents used were Milli-Q water+0.1% formic acid (A) and acetonitrile+0.1% formic acid (B), eluted using the following gradient: 5% B at 0 min; 100% B at 8 min, linear gradient; 8–12 min 100% B; 12–16 min 5% B. The flow rate was 500 μL.min^{-1}. Mass spectrometry settings were as follows: capillary temperature 320°C, source voltage 3.5kV, and sheath gas flow rate 10 L/min. The mass spectrometer operated in both positive and nega-

tive polarity in turn as certain compounds are better ionised in one of the two modes. Mass spectrometric acquisitions were divided into four scan events: positive MS with a window from *m/z* 100 to 1200, followed by three data-dependent MS/MS scans of the three most intense ions detected through the first scan event. Tandem mass spectrometric parameters were set as follows: collision energy=50eV (positive-ion mode) or 45 eV (negative-ion mode), default charge of 1, isolation width of *m/z* 1.3. Dynamic exclusion was disabled. Purine ($C_5H_4N_4$, *m/z* 121.050873), and HP-0921 (hexakis(*1H*, *1H*, *3H*-tetrafluoropropoxy)-phosphazene $C_{18}H_{18}F_{24}N_3O_6P_3$, *m/z* 922.009798 were used as internal lock masses. Full scans were acquired at a resolution of 60.000 (*m/z* 922) and 35.000 (*m/z* 121).

Metabolome data processing and feature identification

I then performed feature identification to convert the raw mass spectrometry data into a list of metabolite features per sample. The following steps were performed separately for the positive and negative ion mode data. I converted raw Agilent .d files to .mzML format using the command line implementation of ProteoWizard release 3.0.23025 (Holman et al. 2014) with default parameters. MS peak detection, chromatogram building, deconvolution, isotope grouping, feature alignment, and gap filling were done in MZMine v3.3.0 (Pluskal et al. 2010). For MS peak detection, I set the noise level threshold to 3000 for MS1 and 20 for MS2 to remove instrumental noise. I performed chromatogram building using the inbuilt ADAP algorithm with a minimum group size of 5 (number of scans). The intensity threshold for chromatogram building was the same as in MS peak detection, and scan-to-scan accuracy was set to 0.001 Da. I used ADAP again to resolve chromatograms, setting S/N at 15 min and feature height to 10. I assigned resolved chromatograms to MS2 features, setting RT tolerance to 0.2 and MS1 to MS2 precursor tolerance to 0.01 Da. I grouped isotopes using the 13C isotope filter with an m/z tolerance of 0.001 Da, RT tolerance of 0.2, maximum charge of one and setting the representative isotope as that with the highest intensity. Grouping isotopes reduces the dataset and results in one metabolite feature per compound. I aligned features between samples using the join aligner with the same values of m/z and RT tolerance as in the previous step. I performed gap-filling using default settings. Gap-filling allowed me to link features without MS2 scans in some samples to features with MS2 scans in others. I removed features with no

related MS2 spectra in any samples. These steps resulted in a presence-absence matrix of metabolite features across all lichen samples. Finally, I exported feature lists for feature-based molecular networking, which is described in the following section.

Molecular networking

I organised and analysed metabolomes via molecular networking. This approach allows for the annotation of compounds by aligning spectra to each other and to curated reference databases based on shared fragmentation patterns (Nothias et al. 2020). I created molecular networks for positive and negative ion mode datasets using Feature-Based Molecular Networking (FBMN) (Nothias et al. 2020) via GNPS (Wang et al. 2016), an open-access online portal for organizing, processing and annotating mass spectrometry data. Within a GNPS molecular network, each metabolite feature is represented as a node and links between nodes represent metabolite features with a fragmentation pattern similarity above a given threshold. Each group of connected nodes is referred to as a subcluster, and nodes within the same subcluster are likely to represent the same or similar compounds. I uploaded the MZMine output files (.mgf and .csv) from the previous section to GNPS (`http://gnps.ucsd.edu`). I set precursor ion mass tolerance and MS/MS fragment ion tolerance to 0.02 Da to cluster spectra across samples and create consensus spectra. I filtered network edges to have a cosine score above 0.55 and more than four matched peaks. All other parameters were left at default values. Spectra in the resulting molecular networks were matched to all public GNPS spectral libraries.

Statistical analyses of metabolomic & genomic data

In this section, I performed statistical analyses on two datasets, the first containing all taxa and the second containing just *Teloschistales*. In an attempt to consolidate positive and negative ion data and remove redundant ions, MolNotator (Olivier-Jimenez et al. 2021) was run with the aid of Dr Damien Olivier-Jimenez (Université de Rennes), developer of the program. MolNotator produced a dataset of computed neutral compounds. I then carried out the metabolomic statistical analysis on five separate datasets: (1) MolNotator computed neutrals, (2) the positive ion MZMine feature list, (3) negative ion MZMine feature list, (4) positive mode GNPS subnetworks and (5) negative mode GNPS subnetworks. Positive and negative ion MZMine and GNPS results were analysed separately as

there is a chance that a single metabolite could be ionised in both modes. Thus combining the presence-absence matrices could lead to redundancy with two columns for the same metabolite feature. Dataset 1 (MolNotator computed neutrals) is the preferred dataset as it allows consolidation of positive and negative data. However, as it is a relatively new approach, I wanted to compare the results to the more traditional MZMine feature lists (Datasets 2 and 3). The GNPS outputs (Datasets 4 and 5) were included as they potentially reduce the complexity of the MZMine feature lists by linking similar compounds into a single subcluster. Furthermore, as GNPS allowed me to annotate metabolite features, I could search for and perform statistical tests on anthraquinone features specifically. I converted all datasets to presence-absence and removed any ions present in the blank samples or, in the case of GNPS datasets, any network subcluster containing a blank ion.

Next, I converted the presence-absence BiG-SCAPE BGC family matrix and the five metabolome matrices into dissimilarity matrices by calculating Jaccard dissimilarity indices via the 'vegdist' function of the R package vegan c.2.5.7. This resulted in six presence-absence matrices, one quantifying BGC similarity between samples, and five quantifying metabolome similarity between samples.

To quantify phylogenetic relatedness between samples, I converted the genome-scale phylogenetic tree into a phylogenetic distance matrix. For the *All Taxa* dataset, I rooted the IQTree maximum likelihood species tree using *Arthoniaceae* as an outgroup. I converted it into a phylogenetic distance matrix using the 'cophenetic.phylo' function of the R package ape v5.6-1 (Paradis et al. 2019). For the *Teloschistales* dataset, I used my 84-taxon time-calibrated MCMCTree phylogeny (see Chapter 3). I removed the *Caliciales* outgroup and rooted it with the *Letrouitineae* clade.

I then used principal coordinates analysis (PCoA) to observe whether BGC and metabolite presence-absence patterns reflected phylogenetic relatedness. I performed two separate PCoAs using vegan's 'wcmdscale' function with the BGC family matrix and the MolNotator computed neutrals dissimilarity matrix as inputs. To understand variation within the *Teloschistales*, I repeated the PCoA excluding all non-*Teloschistales* taxa. Then, to observe which BGC families or metabolite features correlated most strongly with the PCoA datapoint dispersal, I fitted each BGC family or metabolite as an environmental variable on their respective PCoA biplot. I plotted only those variables with a correlation p-value$<$0.01.

I then tested for correlation between the BGC family matrix and each of the five metabolome dissimilarity matrices individually using Mantel tests implemented in vegan. For each of the five BGC-metabolome comparisons, I also performed a partial Mantel test, incorporating the phylogenetic distance matrix to test whether correlations between the BGC family and metabolome matrices were still observed when controlling for phylogenetic relatedness.

To test whether the number of BGCs and metabolites per sample was phylogenetically significant, I performed trait evolution analysis. Since the number of BGCs and metabolites per sample are continuous traits, I calculated Blomberg's K (Blomberg et al. 2003) and Pagel's lambda (Pagel 1999) using the phylosig function of R package phytools (Revell 2012). I then tested the null hypothesis that these trait values (no. BGCs and no. metabolites) show no phylogenetic signal (K or lamba = 0). To do this, I shuffled the trait values on the tree, recalculated K and lambda 10,000 times and then compared these recalculated values to the observed K and lambda values using phylosig. For lambda, an additional step is required where phylosig calculates the likelihood ratio test (LRT) statistic and uses a χ^2 test with 1 degree of freedom to compare the observed and randomized estimates. A significant p-value suggests significant phylogenetic signal in the trait, meaning closely related taxa are more similar than random pairs of taxa. I also tested whether the observed K and lambda differ from those expected under a Brownian model of evolution (K or lambda = 1). For K, I did this by simulating 10,000 datasets under a Brownian model on the species tree using the phytools function fastBM and then calculating K. I then counted how many times simulated values were below the observed K. For lambda, I calculated the likelihood that lambda = 1 and then compared this with the maximum likelihood estimate of lambda I obtained from my dataset. Again, I calculated the LRT statistic and used a χ^2 test with one d.f.

For each continuous trait, I also compared the fit of three models of trait evolution: Brownian motion (Felsenstein 1973), early burst model (Harmon et al. 2010) and Ornstein-Uhlenbeck model (Butler et al. 2004). Brownian motion describes a scenario where traits evolve stochastically along the phylogeny, with changes occurring randomly and independently. The early burst model instead describes a scenario where most phenotypic divergence occurs early in the phylogeny, which is often linked to adaptive radiations. Finally, Ornstein-Uhlenbeck describes a scenario where phenotypes converge to one or more

stable adaptive peaks. I used the phytools function fitContinuous to fit the models and then used AIC and weighted AIC to determine which fit best.

Finally, I tested whether the presence/absence patterns of individual BGC families or metabolites were consistent with the species tree (IQTree ML phylogeny for *All Taxa* and 84-taxon MCMCTree phylogeny for *Teloschistales*). For this, I calculated Fritz and Purvis' D-statistic for each individual BGC family and metabolite using the phylo.d function of the R package caper (Orme et al. 2012). The function also calculates the probability that the observed D value significantly differs from a D value expected under both a random (D = 1) and Brownian motion (D = 0) model of evolution across the tree.

Mantel tests, Blomberg's K and Pagel's lambda, and Fritz and Purvis' D-statistic analyses were all performed for both the *All Taxa* dataset and the pruned *Teloschistales* dataset.

Ancestral state reconstruction and evolutionary modelling of anthraquinone BGCs and metabolites

To understand how gains and losses at the gene level may affect BGC and metabolite evolution, I then focused on the *Teloschistales* anthraquinone BGCs characterised in Chapter 2 and the annotated anthraquinone GNPS subnetworks as a case study. First, I used ancestral state reconstruction at the gene level to understand how anthraquinone BGCs have evolved. I then performed an enrichment analysis of anthraquinone GNPS subnetworks and joint modelling of BGCs and metabolites to test whether BGC evolution affects the diversity of anthraquinones produced.

For the ancestral state reconstruction, I first identified the anthraquinone BGC by looking for BiG-SCAPE subnetworks containing BGCs with the 4-gene BGC structure (EthD, PKSI, MβL, ABC) characterised in Chapter 2 and containing known anthraquinone BGCs from the MIBiG reference dataset. I excluded BGCs with less than three genes as these are likely a result of fragmented assemblies. I reconstructed the ancestral state of each gene within the anthraquinone BGC individually. For this, I focused on genes present in at least three taxa. This resulted in a dataset of ten genes which I have named based on their annotated Pfam domains: gamma-glutamyltransferase, kinase phosphorylation, Domain Unknown Function (DUF), EthD, PKSI, MβL, ABC transporter, methyltransferase, amino acid permease and oxidoreductase.

I identified homology between each gene within the BGC across taxa using cblaster (Gilchrist et al. 2021). For this, I used the BGCs from *Teloschistes flavicans* (LIQ166TFLAV), *Xanthocarpia ochracea* (LIQ230CAOCH) and *Caloplaca* s.l. sp 3 (LIQ195CASP) as queries. These taxa covered the three *Teloschistales* subfamilies and had anthraquinones on long contigs, a good proxy for a high-quality BGC annotation. For each gene within the anthraquinone BGC, I checked if the sequences from different taxa were homologous and fell into a single orthogroup using Orthofinder's 'Orthogroups.tsv' file.

For each gene, I assigned the following trait values for each species: absent = gene is completely absent from the genome, clustered = gene is present within the anthraquinone BGC, unclustered = gene is absent from the anthraquinone BGC but a homolog is present elsewhere in the genome. Therefore, each gene was considered a discrete trait with three possible states (absent, clustered and unclustered). Before running ancestral reconstruction, I compared the fit of the equal rates (ER) and all-rates different (ARD) model for each gene using the phytools fitMK function and AIC values. The ARD model was a better fit for 8/10 genes; thus I chose this model for ancestral state reconstruction. I performed joint and marginal ancestral character state reconstructions for each gene with the ARD model via the corHMM R function (Beaulieu et al. 2013). corHMM was chosen as it allows for multiple evolutionary models and both joint and marginal reconstruction. Finally, I plotted the posterior probabilities of each character state from the marginal estimation on the nodes of the phylogeny.

Next, to test whether anthraquinone BGC evolution leads to the diversification of anthraquinone metabolite profiles, I performed an enrichment analysis. First, I filtered the GNPS network for any subnetworks linked to anthraquinones in the GNPS spectral libraries. I tested whether *Teloschistales* samples were over-represented in GNPS anthraquinone subnetworks using hypergeometric enrichment tests via the function 'phyper' in R 4.3.0 (R Development Core Team 2019). For each lichen sample, the test requires four parameters: the number of nodes within all anthraquinone subnetworks that the sample was detected in (q), the number of nodes within all subnetworks of the network that the sample was detected in (m), the number of nodes within all subnetworks of the network that the sample was NOT detected in (n), and the total number of nodes in all anthraquinone subnetworks combined (k). This produced a p-value for each species

showing whether that species was significantly enriched in anthraquinone subnetworks.

Finally, I tested whether the evolution of anthraquinone diversity differs between *Teloschistales* taxa and outgroups and whether anthraquinone evolution correlates with the evolution of individual genes with the anthraquinone BGC. Anthraquinone diversity was measured as the number of anthraquinone-like metabolite features detected in the GNPS network per species (the same as parameter q in enrichment analysis). I compared the rate of anthraquinone evolution between *Teloschistales* taxa and outgroups using the R package OUwie v2.12 (Beaulieu et al. 2012). Treating the number of anthraquinones per sample as a continuous character, I compared three models of continuous trait evolution: (1) Brownian motion model with a single rate of evolution (σ^2) across the whole phylogeny, (2) Brownian motion with separate evolutionary rates for *Teloschistales* and outgroups, and (3) Ornstein-Uhlenbeck (OU) with separate optimal trait values (θ) for *Teloschistales* and outgroups. I compared the fit of the three models using weighted AIC values.

To test whether anthraquinone evolution correlates with anthraquinone BGC evolution, I jointly modelled the evolution of the number of anthraquinone-like metabolite features (continuous trait) with the evolution of each gene within the anthraquinone BGC (discrete trait). For this, I used the *hOUwie* function within OUwie, which explicitly estimates the joint likelihood for discrete and continuous characters in a single evolutionary framework (Boyko et al. 2023). Tip states for each gene were coded with three states (absent, clustered and unclustered) as in the ancestral state reconstruction. I used the ancestral states reconstructed by corHMM to map gene trait values to internal nodes of the phylogeny. I used the ARD model for discrete gene traits and compared four models for the continuous anthraquinone trait:

1. **BM1** Brownian motion model with no link between anthraquinone number evolution and discrete gene character evolution, i.e. all parameters are identical for states (absent), (clustered) and (unclustered)

2. **BMV** Brownian motion where rates of anthraquinone number evolution (σ^2) differ between state (absent), (clustered) and (unclustered)

3. **OU1** Ornstein-Uhlenbeck with no link between anthraquinone number evolution and discrete gene character evolution

4. **OUM** Ornstein-Uhlenbeck where optimal trait values of anthraquinone number evolution (θ) differ between state (absent), (clustered) and (unclustered)

For each gene, I fitted the four *hOUwie* models with 100 stochastic mappings per iteration. I determined the best fit continuous model using weighted BIC.

4.4 Results

Genome data

To understand BGC evolution broadly in lichen-forming fungi and disentangle clade-specific patterns, I sampled and sequenced eight new non-*Teloschistales* taxa. Supplementary Table A.8 shows novel mycobiont genome assembly statistics. Assembly sizes ranged from 24.77 Mbp to 33.79 Mbp. BUSCO completeness ranged from 83.70% to 92.67%, and duplicated BUSCOs were all below 1.5% (mean 0.20%). Predicted genes ranged from 6514 to 9778. New genomes were combined with 74 mycobiont genomes from previous chapters for a total of 82 genomes used for biosynthetic gene cluster annotation and analysis. BGC annotation identified 3400 BGCs across the 82 genomes. The number of BGCs per genome ranged from 18 to 73 (mean = 41). BiG-SCAPE network analysis (Navarro-Muñoz et al. 2020) grouped the BGCs into 1362 biosynthetic gene cluster families (BGC families), of which 25 contained MIBiG reference BGCs.

Metabolome data

Untargeted metabolomes were successfully generated for 79 taxa. Heatmaps showing the presence-absence of MolNotator computed neutral metabolites are shown in Figure 4.1. In total, 476 unique computed neutral metabolites were detected, of which 397 were detected in two or more taxa. Clustering samples based on their metabolite phenotype similarity (left heatmap and tree Figure 4.1) does not match the clustering of samples based on the ML phylogeny (right heatmap and tree Figure 4.1). Only three groups of samples are equally recovered by both methods: the grouping of *Coniocarpon cinnabarinum* (accessions: LIQ171ARSP, LIQ211ARSP), the *Letrouitia transgressa* group (accessions: LIQ93LETR, LIQ94LETR-2), and the *Wetmoreana* group (accessions: LIQ185CASP, LIQ186CAVE, LIQ189CABR). This suggests metabolome similarity does not reflect phylogenetic relatedness. A subsection of metabolites was shared across most samples (see

clade highlighted with '*' in Figure 4.1). However, most detected metabolites were found only in a handful of samples.

GNPS feature-based molecular networking produced two molecular networks for positive and negative ion mode. The positive network contained 5697 nodes, 6768 edges, and 2258 connected components. The negative network contained 1788 nodes, 2127 edges, and 657 connected components. GNPS was able to annotate a small number of the detected metabolites. For positive mode, 22 dereplicated ions were matched to 11 known metabolites. These were B27A19, caloploicin, canarione, isokaempferide, emodin, parietin, pheophorbide A, secalonic acid D, secalonic acid F, usnic acid and variolaric acid. For negative mode, 22 dereplicated ions were matched to 17 known metabolites. These were 2-chloro-1,3,8-trihydroxy-6-(hydroxymethyl)anthracene-9,10-dione, allo-pertusaric acid, caloploicin, citreorosein, confluentic acid, divaricatic acid, endocrocin, gyrophoric acid, isokaempferide, isoskyrin, lecanoric acid, 2-(((2-Ethylhexyl)oxy)carbonyl)benzoic acid, nephromopsic acid, ovoic acid, pseudoplacodiolic acid, sekikaic acid and variolaric acid. Patterns of annotated mass features are shown in Supplementary Figure A.11. Some mass features were found in most samples, such as emodin ($n = 63$), parietin ($n = 70$), citreorosein ($n = 52$) and pheorphobide A ($n = 64$). The vast majority of mass features could not be annotated, however, likely due to the lack of reference lichen metabolites in public databases.

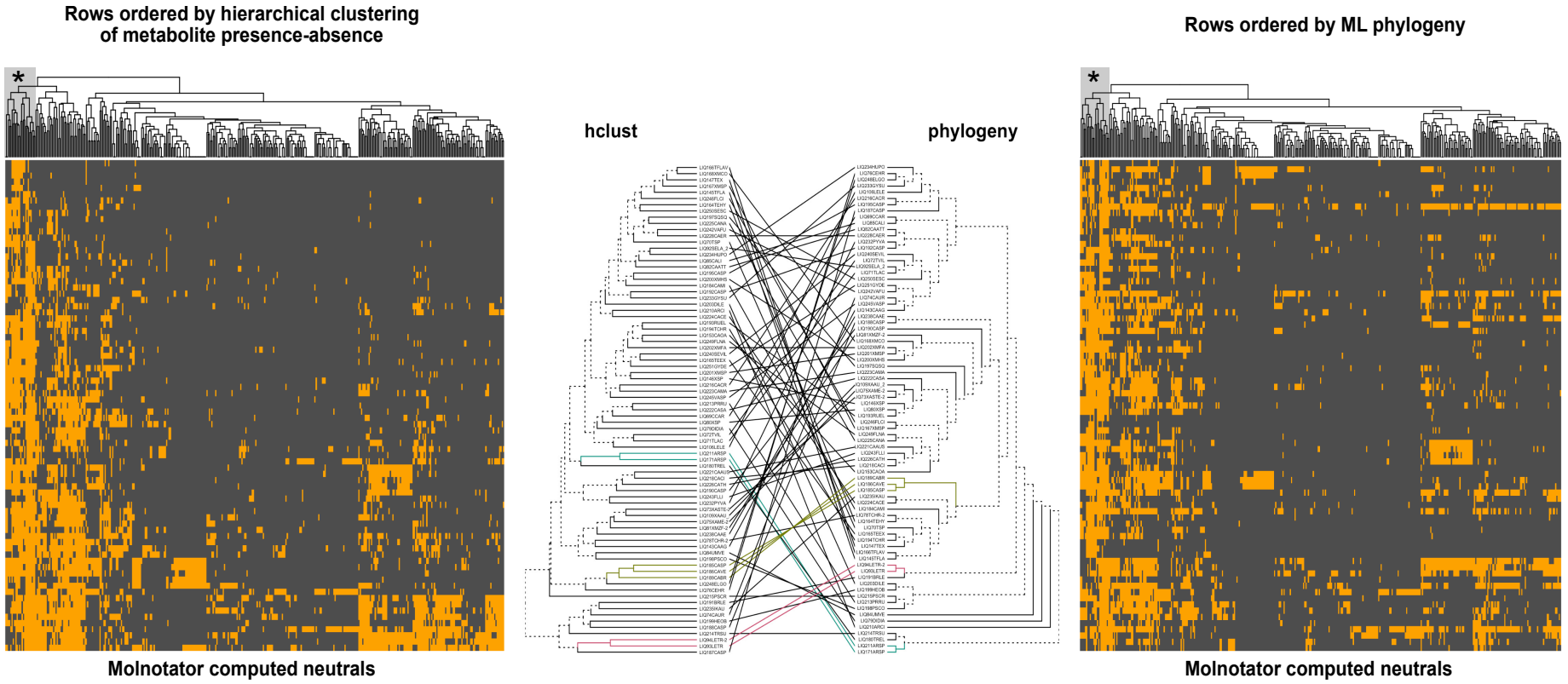


Figure 4.1: Untargeted metabolomics analysis of 79 lichen-forming fungi. The presence-absence matrices on the left and right represent the same metabolite dataset with lichen taxa (rows) ordered by either hierarchical clustering of metabolite presence-absence (left tree) or maximum likelihood phylogeny of 1351 genes (right tree). Columns of heatmaps represent computed neutral metabolites from MolNotator (n = 397 with singletons removed). Dendrograms above heatmaps show hierarchical clustering of metabolites from presence-absence patterns. Clade labelled with asterisk shows subsection of metabolites shared across most taxa. Presence = orange, absence = grey. Central tanglegram highlights differences between the hierarchical clustering dendrogram of samples from metabolite presence-absence (hclust) and the ML phylogeny. Solid colour branches highlight clades found in both trees. Central lines connect the same lichen sample in both trees

Paired metabolome-genome analysis

Clustering analysis of the BGC family and metabolite (computed neutrals) presence-absence variation for the full dataset showed that most of the variation occurred between *Teloschistales* taxa. Non-*Teloschistales* variation was low and fell within the variation of *Teloschistales* clade on the PCoA biplots (Supplementary Figure A.12). Given that this result is most likely due to sampling bias towards the *Teloschistales*, I decided to focus on within-order BGC and metabolite variation for subsequent results.

Within *Teloschistales*, PCoA analysis of BGC family and secondary metabolite presence-absence variation showed remarkably different patterns (Figure 4.2). BGC data (left-hand plot Figure 4.2) showed that samples clustered strongly according to taxonomic subfamily. Only the outgroup *Letrouitineae* fell in the centre of the other clusters. Fitting each BGC family as an environmental variable on the BGC plot showed this separation was driven by each *Teloschistaceae* subfamily possessing a unique set of BGCs (Supplementary Figure A.13). The computed neutrals metabolite clustering, on the other hand, showed no grouping according to subfamilies, with most datapoints intermixed (right-hand plot Figure 4.2). A handful of outliers are seen belonging to *Teloschistes flavicans* (accessions LIQ145TFLA, LIQ166TFLAV) *T. exilis* (LIQ147TEX), *Xanthomendoza* sp. (LIQ167XMSP), and *X. concinna* (LIQ168XMCO).

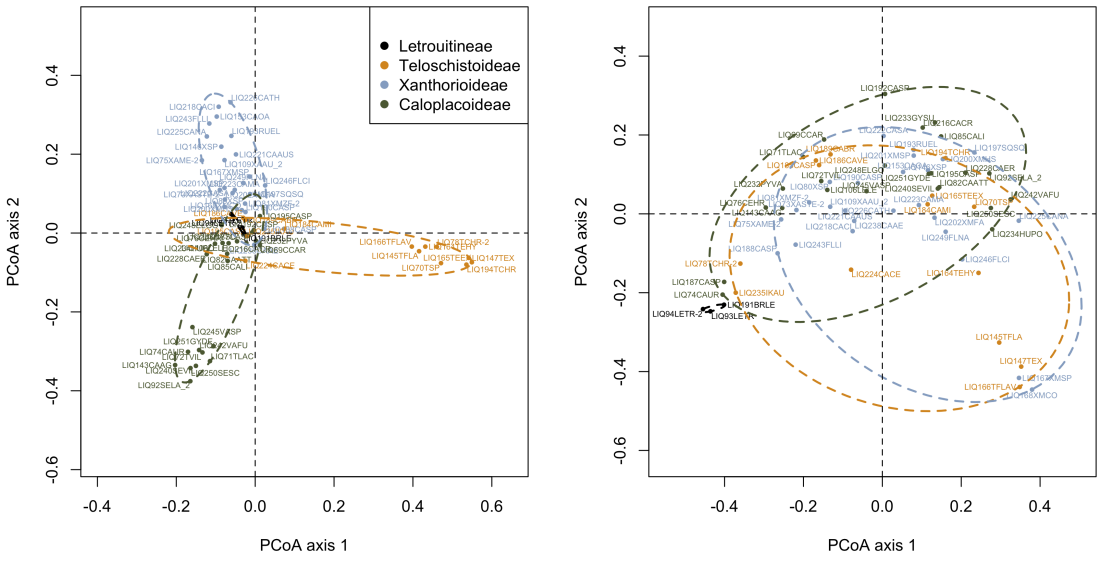


Figure 4.2: PCoA plots of BGC family (left) and computed neutral metabolites (right) presence-absence patterns in the *Teloschistales*. Colours represent taxonomic subfamilies of *Teloschistaceae* and outgroup suborder *Letrouitineae*, with ellipses drawn around groups.

Mantel tests of the BGC family and metabolite presence-absence patterns showed that distance matrices of all five metabolite datasets (computed neutrals, +ve ions, -ve ions, +ve GNPS subnetworks, -ve GNPS subnetworks) significantly correlated with the BGC family distance matrix. However, subsequent partial Mantel tests showed that after accounting for phylogenetic distance, metabolite and BGCs distance no longer correlate for *Teloschistales* (Table 4.1).

Table 4.1: Mantel test statistics based on Pearson's product-moment correlation testing correlation between BGC family presence-absence distance matrix and metabolite distance matrices. Partial Mantel statistic accounts for phylogenetic distance. Significant p-values are marked with *.

Dataset	Trait	Mantel statistic (r)	p-value	Partial Mantel statistic (r)	p-value
All Taxa	neutrals	0.2013	0.001*	0.08851	0.002*
	+ve ions	0.1663	0.001*	0.07129	0.007*
	-ve ions	0.2825	0.001*	0.09066	0.001*
	+ve GNPS	0.1447	0.001*	0.0636	0.012*
	-ve GNPS	0.2636	0.001*	0.07958	0.006*
Teloschistales	neutrals	0.1659	0.001*	0.02547	0.215
	+ve ions	0.1305	0.001*	0.04071	0.141
	-ve ions	0.2044	0.001*	-0.0313	0.767
	+ve GNPS	0.1177	0.001*	0.04718	0.125
	-ve GNPS	0.1879	0.001*	-0.02394	0.722

Blomberg's K and Pagel's lambda analyses suggest that the number of BGCs and metabolites per sample correlates with the lichen phylogeny (i.e. K and lambda significantly differ from 0) for both the *All Taxa* dataset and the *Teloschistales*-only dataset (Table 4.2). Therefore, closely related taxa are more likely to have similar numbers of BGCs and metabolites than random pairs of taxa. For all traits in the *All Taxa* dataset, I reject the null hypothesis that they follow a Brownian motion model of evolution (i.e. K and lambda significantly differ from 1). For the *Teloschistales* dataset, Blomberg's K suggests that the number of BGCs, negative ions and negative GNPS subnetworks are consistent with Brownian motion. However, this is not supported by Pagel's lambda. Comparing the fit of three alternative evolutionary models shows the Orstein-Uhlenbeck model is the best fit for all traits in both the *All Taxa* and *Teloschistales* dataset (Supplementary Table A.9).

Fritz and Purvis' D statistic analysis shows that the distributions of 93% of BGC families are consistent with Brownian motion across the phylogeny ($D = 0$, $p > 0.05$), and only 7% follow a random distribution ($D = 1$, $p < 0.05$). For computed neutral metabolites, 72% are consistent with Brownian motion and 22% are randomly distributed (Figure 4.3).

The same pattern is seen for both the *All Taxa* and *Teloschistales* datasets. Therefore a slightly higher proportion of BGCs follow Brownian motion compared to metabolites.

Table 4.2: Results of Blomberg's K and Pagel lambda (λ) analysis for number of BGCs, computed neutrals, positive ions, negative ions, positive GNPS subnetworks and negative GNPS subnetworks. Statistical tests for each trait are grouped by hypotheses. A K or λ significantly different from 0 (K or $\lambda = 0$) suggests trait shows phylogenetic signal. A K or λ not significantly different from 1 (K or $\lambda = 1$) suggests trait follows Brownian motion (marked with †).

				Phylogenetic signal		**Brownian motion**	
Dataset	**Trait**	**K**	λ	**K=0 p-value**	**λ=0 p-value**	**K=1 p-value**	**λ=1 p-value**
All Taxa	bgcs	0.1541	0.9507	0.0001	<0.0001	<0.0001	0.0001
	neutrals	0.1423	0.6507	0.0002	0.0006	<0.0001	<0.0001
	+ve ions	0.0954	0.4771	0.0019	0.0264	<0.0001	<0.0001
	-ve ions	0.2037	0.8949	0.0001	<0.0001	0.0012	<0.0001
	+ve GNPS	0.093	0.464	0.0016	0.0344	<0.0001	<0.0001
	-ve GNPS	0.1828	0.8111	0.0001	<0.0001	0.0003	<0.0001
Teloschistales	bgcs	0.4932	0.8954	0.0001	<0.0001	0.0511†	0.0006
	neutrals	0.4125	0.7557	0.0003	0.0005	0.0142	<0.0001
	+ve ions	0.3261	0.6398	0.0007	0.0119	0.0008	<0.0001
	-ve ions	0.6552	0.9095	0.0001	<0.0001	0.2260†	0.0010
	+ve GNPS	0.3223	0.6167	0.0005	0.0164	0.0007	<0.0001
	-ve GNPS	0.6116	0.8934	0.0001	<0.0001	0.1701†	0.0004

Evolution of anthraquinone metabolites and BGCs

After assessing patterns of variation and evolution for all secondary metabolites and associated BGCs in this group of lichens, I focused exclusively on anthraquinones to test whether patterns relating to key metabolites reflect global metabolome/genome patterns. GNPS molecular networking revealed nine dereplicated ions matching the anthraquinones emodin, parietin, canarione, citreosein, 7-chlorocitreosein, endocrocin and isoskyrin (Figure 4.4). These were not restricted to the *Teloschistales*. Dereplicated nodes were connected to 175 additional nodes in the GNPS networks. As GNPS connects nodes based on similar fragmentation patterns, these additional nodes may represent other anthraquinones. I performed enrichment analysis to see if *Teloschistales* samples were enriched in anthraquinone GNPS subnetworks. This showed that apart from a few exceptions, *Teloschistales* taxa, in general, are not more enriched in anthraquinone subnetworks compared to other metabolite group subnetworks (Supplementary Figure A.14). This suggests *Teloschistales* are not more enriched in anthraquinone diversity than outgroup taxa.

BiG-SCAPE analysis revealed six putative anthraquinone BGC families. Four of these

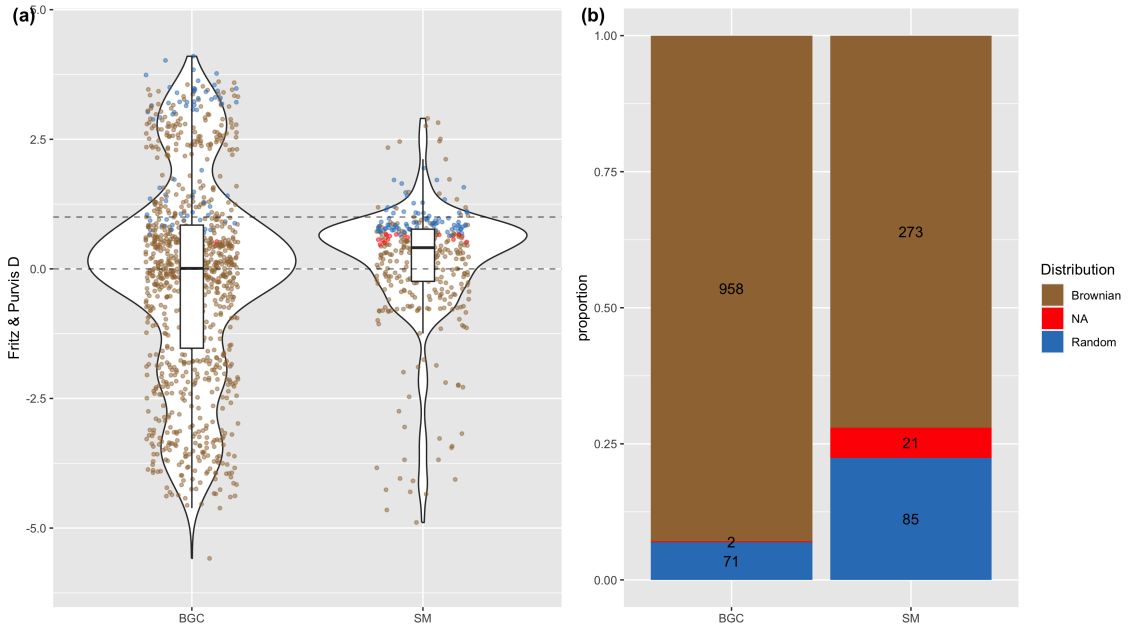


Figure 4.3: Fritz and Purvis' D-statistics of Biosynthetic Gene Clusters (BGC) families and Secondary Metabolites (SMs) for the 79-taxon *All Taxa* dataset. (**a**) Violin and boxplot showing D-statistics of BGC families and SMs. Individual D-statistics are shown with data points. The colour shows whether D-statistic significantly differs from 0 (Brownian distribution - brown) or 1 (Random distribution - blue). Red points indicate BGC families/SMs where the statistical test could not determine their distribution. (**b**) Bar chart showing the proportions of BGC families and SMs consistent with Brownian motion and Random distribution. Numbers within bars indicate counts within each category.

showed the typical four-gene anthraquinone structure characterised in Chapter 2. These were restricted to *Teloschistales* and two *Psoraceae* genomes. Two BGC families, largely restricted to outgroup taxa, showed BGCs without the ABC transporter. In addition to the four core genes consistently found in Chapter 2 (EthD, PKSI, MβL, ABC transporter), six additional genes were consistently found in the anthraquinone BGCs of at least three species. These genes encoded the protein domains methyltransferase, kinase phosphorylation, gamma-glutamyltransferase, amino acid permease, oxidoreductase and Domain Unknown Function (DUF). There was a high level of synteny between anthraquinone BGCs in different *Teloschistales* taxa, especially for the four core genes and the methyltransferase. The organisation of these genes is summarised in Figure 4.5.

Two additional BGC families were linked to the typical anthraquinone BGC families in the BiG-SCAPE network. These BGCs had markedly different structures and gene content and were found in non-*Teloschistales* taxa and a handful of *Teloschistales*. Many of these BGCs contained halogenase genes suggesting they may be involved in halogenated anthraquinone biosynthesis. For *Teloschistales* taxa, these halogenase-type anthraquinone BGCs were found in addition to the typical anthraquinone BGC rather than instead of.

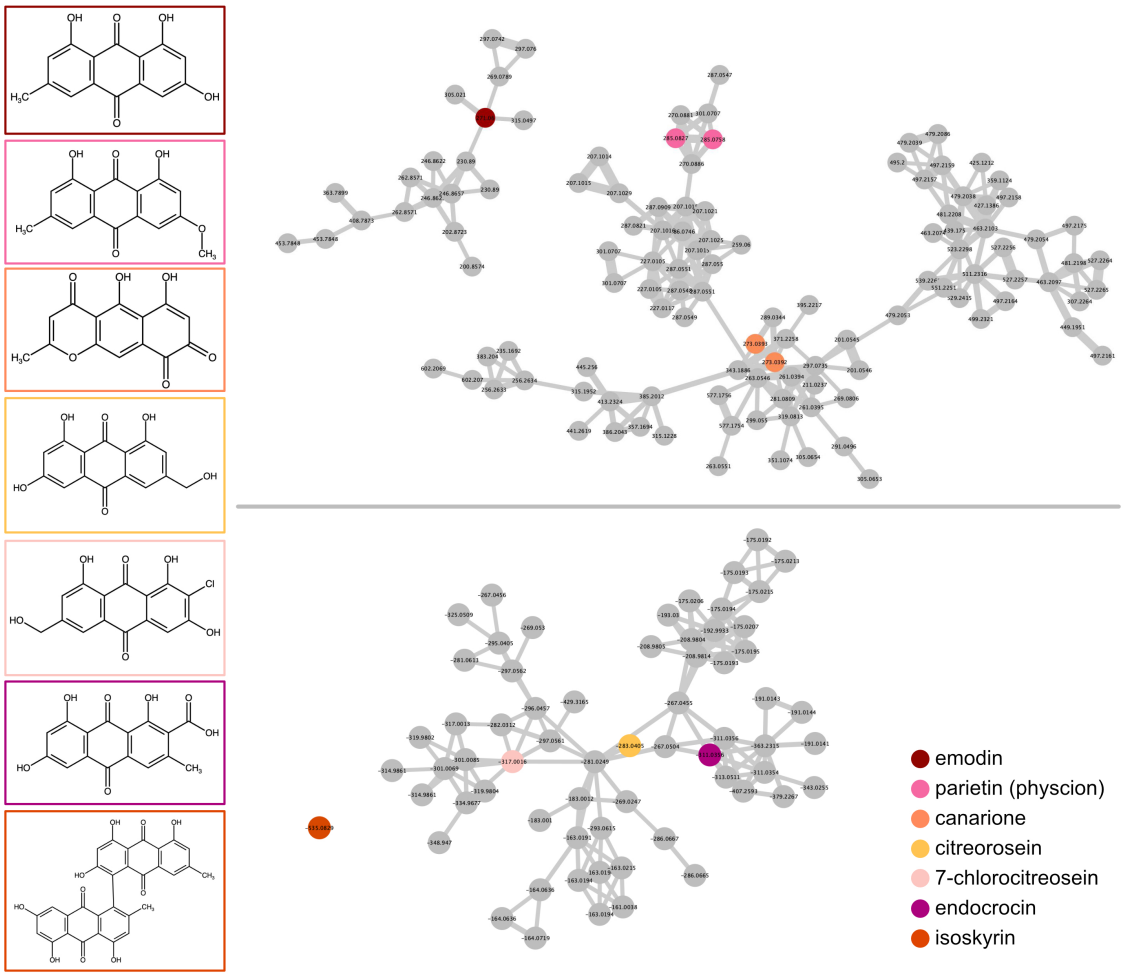


Figure 4.4: Anthraquinone subnetworks from GNPS feature-based molecular networking. Nodes represent detected metabolite features, annotated with their mass. Edges connect features with a cosine similarity score >0.55, edge width represents cosine score. Coloured nodes highlight those matching compounds in the GNPS MS^2 spectral libraries. Matching compound chemical structures are shown on the left-hand side. subnetworks above and below the grey line are from positive and negative mode GNPS analysis, respectively. Network visualised in Cytoscape v3.9.1.

Therefore they do not appear homologous to the typical *Teloschistales*-type BGCs and were not included in the ancestral state reconstruction.

Ancestral state reconstruction suggested that a five-gene anthraquinone BGC (4 core genes + methyltransferase) was present in the common ancestor of the *Teloschistales* and a three-gene BGC may have been present in the common ancestor of all taxa included (the ancestor of classes *Arthoniomycetes*, *Eurotiomycetes* and *Lecanoromycetes*) (Figure 4.6). Reconstruction also suggests that, apart from the ABC transporter, homologs of the remaining genes were present but unclustered in the genome of the common ancestor of all taxa. Ancestral state reconstruction, therefore, suggests a potential scenario of anthraquinone BGC evolution in the *Teloschistales*; constituent genes were scattered across the genome in the ancestor of distantly related lichen-forming fungi in the *Ascomycota*.

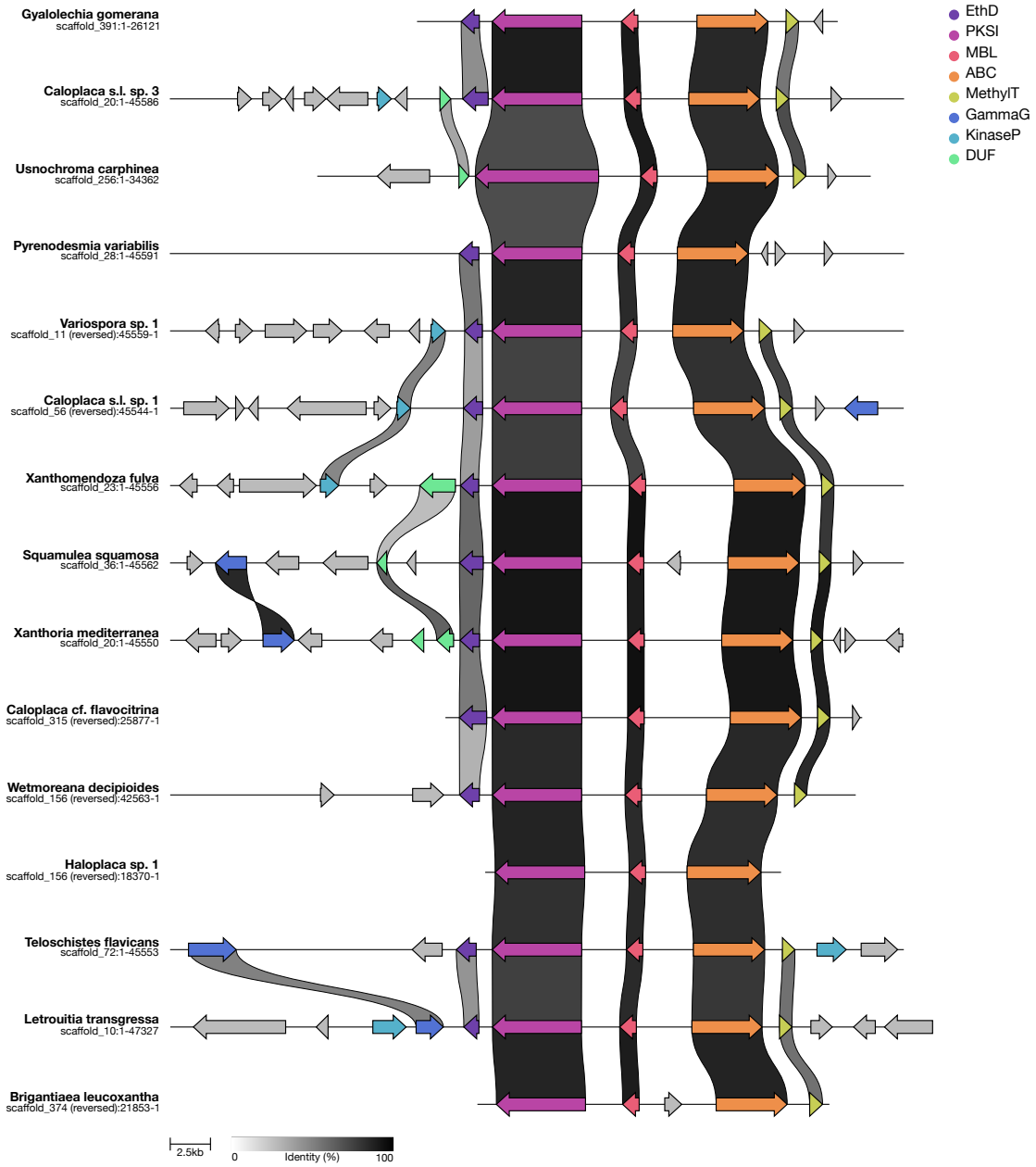


Figure 4.5: Synteny plot showing the organization of anthraquinone clusters in 15 representative *Teloschistales* species. Genomic coordinates of BGCs are shown under taxon names. Gene colours represent homologous genes as determined by all vs. all blasts. Only genes present in three or more taxa are coloured. The darkness of blocks linking genes represents per cent sequence identity between genes. Only links >30% sequence identity are shown. Taxa were chosen to represent all major clades of *Teloschistales* and are ordered phylogenetically.

Genome rearrangements subsequently brought together EthD, PKSI, MβL, ABC transporter and Methyltransferase genes in the ancestor of *Teloschistales* order. Following this, further gains of accessory genes occurred within the order. For example, the *Teloschistes* genus appears to have gained a kinase phosphorylation gene, and the *Xanthoriodeae* gained a DUF gene. Some clades show secondary losses, such as the loss of a methyltransferase within *Pyrenodesmia*.

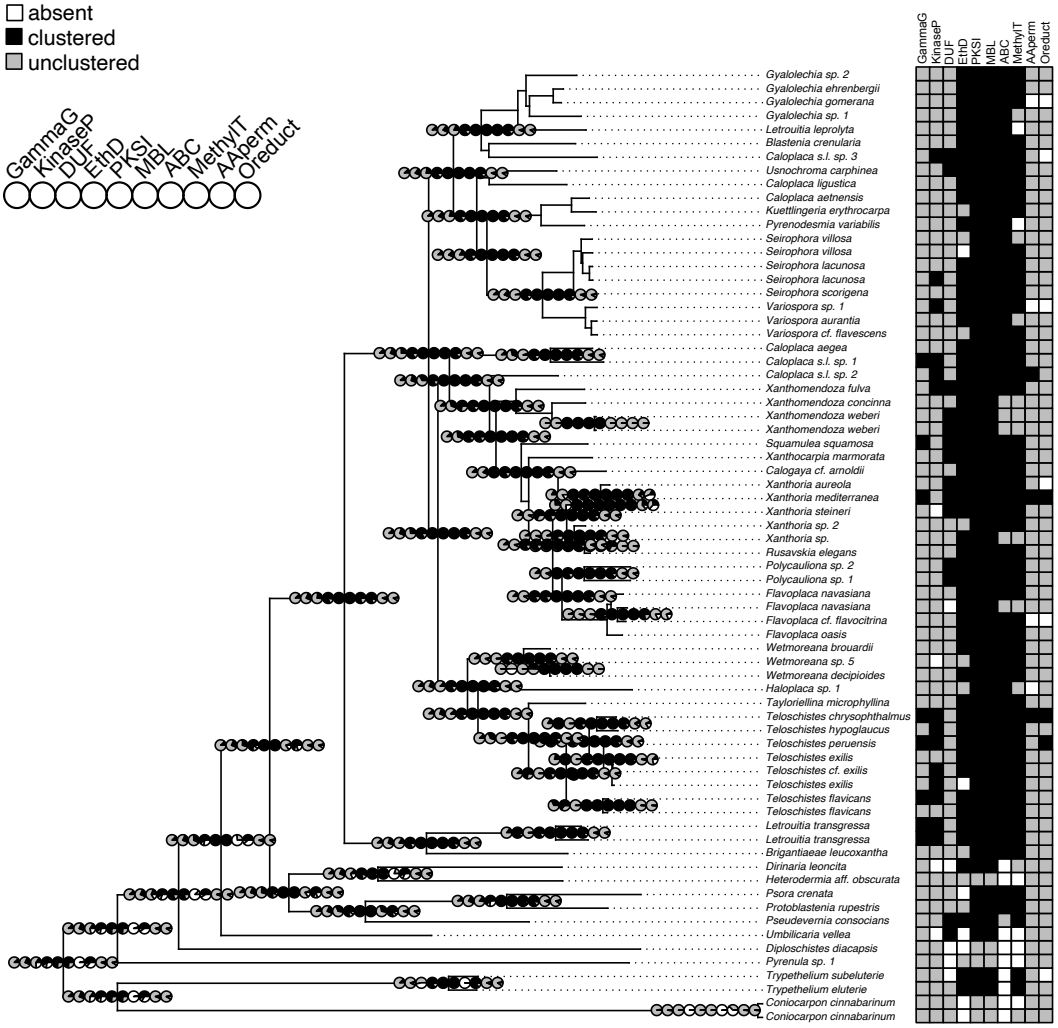


Figure 4.6: Ancestral state reconstruction of genes within anthraquinone BGC. Maximum Likelihood phylogenetic tree reconstructed with IQTree is depicted. Presence-absence matrix of each gene included in the analysis is shown on the right. Posterior probabilities of each character state from the ancestral state reconstruction marginal estimation are shown as pie charts on the nodes of the phylogeny. Some internal nodes with the same state as ancestor are not shown for readability. White = gene absent from genome (absent), black = gene present in anthraquinone BGC (clustered), grey = gene absent from anthraquinone BGC but present elsewhere in the genome (unclustered).

Modelling anthraquinone evolution across the phylogeny, as measured by the number of GNPS anthraquinone-like features per sample, showed no difference in the evolutionary rate of anthraquinones between *Teloschistales* taxa and outgroups. The best-fit model was a Brownian motion model with a single rate of evolution (σ^2) across the whole phylogeny (BM single rate AIC = 1.41×10^{11}, BM separate rates AIC = 2.81×10^2, OU separate optima AIC = 2.69×10^2).

To test for the correlation between the evolution of anthraquinone number (continuous trait) and the evolution of each gene within the anthraquinone BGC (discrete trait),

I fitted four *hOUwie* models. Results showed that the best-fitting model for all genes was a Brownian motion model with no difference in parameters for states (absent), (clustered) and (unclustered) (Supplementary Table A.10). This suggests that the evolution of individual genes within the anthraquinone BGC does not correlate with the evolution of anthraquinone diversity.

4.5 Discussion

In this chapter, I have presented paired genome-metabolome analysis for a diverse clade of lichen-forming fungi using metagenomics and untargeted LCMS/MS metabolomics. I found a complex relationship between variation at the genomic level and variation at the metabolome level. When looking at the BGC repertoires as a whole, lichen samples cluster strongly according to phylogeny, whereas for metabolomes, samples show no clear phylogenetic clustering. However, when looking at presence-absence patterns of individual metabolites, a high proportion correlate with the phylogeny. Furthermore, the overall number of both BGCs and metabolites per sample is consistent with phylogeny suggesting that closely related taxa tend to have similar-sized metabolomes. I detected a set of mass features representing potential anthraquinones, many of which could not be linked to compounds in public datasets. When describing their associated BGCs I observed a high level of synteny between anthraquinone BGCs in different *Teloschistales* taxa. Ancestral state reconstruction of these BGCs in the *Teloschistales* suggested a scenario of BGC evolution consistent with a shuffling of existing genes combined with novel accessory gene gains. However, diversification at the gene cluster level does not correlate with the diversity of anthraquinones produced.

Given that the lichen taxa studied show highly structured, divergent BGCs, it is surprising how much their metabolite profiles overlap (at least according to the clustering analysis shown in Figure 4.2). If this is true, what is the selective advantage of diversifying the BGC repertoire if the secondary metabolite profiles remain relatively stable? Multiple factors may contribute to the lack of a one-to-one relationship between BGCs and metabolites. Firstly, a single BGC may produce different compounds under different conditions or gene expression regimes. This has been shown in both lichen-forming (Singh et al. 2021b) and non-lichen-forming fungi (Wasil et al. 2013). This may also explain why

lichen secondary metabolism varies extensively under different growth conditions, be it experimental lab conditions (Brunauer et al. 2007) or field conditions (Singh et al. 2021a). The fact that lichens retain the genomic capacity to produce many additional metabolites even if they are not normally synthesised suggests that a diverse standing BGC repertoire is advantageous. This may give lichens a more adaptive metabolome to cope with changing climatic conditions or ecological niches.

In contrast to one BGC producing different compounds, different unrelated BGCs can converge to produce the same compound. Convergent compound evolution is widespread within plants (Pichersky et al. 2011) and bacteria (Chevrette et al. 2020) and there are even examples of cross-kingdom convergence. Therefore the lack of clustering and phylogenetic signal seen in metabolomes as a whole may reflect the fact that whilst many compounds are shared among taxa, the BGCs underpinning them are different. This is observed here for anthraquinones and will be discussed in more detail following the discussion of the ancestral state reconstruction.

Another factor that may disrupt a direct one-to-one relationship between metabolome and genome is the presence of non-target metabolites in the samples. These could be from non-mycobiont symbionts, for example, the chlorophyll derivative phaeophorbide A, which is likely from the photobiont. Although it may add some noise to the data, I do not believe this would majorly impact results as lichen secondary metabolites are predominantly produced by the mycobiont (Gill et al. 2023). Other non-target metabolites, such as primary metabolites or intermediate products, could also have an impact. The acetone extraction protocol used here should filter out primary metabolites. However, for intermediate products, without detailed knowledge of biosynthetic pathways, it is difficult to separate these from end products. Whilst we have this knowledge for some key lichen metabolites (Stocker-Wörgötter 2008), most intermediates are not in the reference spectral databases and, therefore cannot be easily annotated using the approach implemented here. Furthermore, compounds such as emodin which are common intermediates in non-lichenised fungi appear to be end products in lichens.

A final complication is that the metabolome data in this study is presence-absence and not abundance. Therefore it is possible that the genomic variation observed leads to changes in the relative abundance of metabolites. Different proportional combinations of metabolites are commonly observed in lichen-forming fungi (so-called ‘chemosyndromes’)

(Culberson et al. 1976). Within the *Teloschistaceae*, chemosyndromes seem to be primarily stable at species level, with no consistent patterns at higher taxonomic levels. For example, HPLC analysis by Arup et al. (2013) showed that whilst the entire *Xanthorioideae* subfamily displays an almost uniform chemosyndrome, the *Caloplacoideae* is much more variable. This is not observed here, with *Xanthorioideae* and *Caloplacoideae* showing similar levels of metabolome variation (Figure 4.2). However, repeating this with abundance data may tell a different story.

To understand patterns of trait diversity, it is crucial to consider the evolutionary process underpinning them. Selection acts on the phenotype of a trait, which for anthraquinones would be UV protection and potentially also defence against microbes or grazers. As compound structure determines biological activity and function, there may be a stronger selective pressure to maintain chemical structure than BGC structure, especially in accessory gene regions. This may explain the higher variation in BGC repertoires compared to metabolomes. Whilst UV protection may have initially driven pigment evolution, once the compound evolved, there would be no benefit of further diversifying its structure. Therefore, UV protection does not seem a likely cause of the highly diverse anthraquinone metabolite and BGC repertoires we observe here (Figures 4.4 and 4.5). Coevolution with antagonistic symbionts may provide a more plausible explanation. Lichens host diverse and variable microbial communities, including host specialist lichen parasites like lichenicolous fungi. As parasites adapt to the antimicrobial defences of the lichen host, diversifying the secondary metabolome becomes advantageous. This evolutionary arms race between mycobionts and antagonistic symbionts may result in 'Red Queen' dynamics, providing the driving force behind persistent metabolite diversification. As axenic culture methods develop, controlled evolution co-culture experiments could explore this hypothesis further by testing if constant antagonistic coevolution with parasites promotes mycobiont BGC and metabolite diversification.

Interestingly, despite the PCoA showing a lack of phylogenetic signal in metabolome dissimilarity, Blomberg's K and Pagel's lambda suggest that the overall number of metabolites and BGCs is consistent with the phylogeny. This means closely related taxa are more likely to have similar-sized metabolite and BGC repertoires than distantly related taxa. The model comparison showed that the Ornstein-Uhlenbeck model is the best fit for the data. This is described as a stochastic process across the phylogeny that tends to revert to

an optimal value (Revell et al. 2022). Therefore although BGC and metabolite turnover may be high between closely related species, there may still be a tendency to revert to similar numbers of BGCs or metabolites. This indicates physiological limitations on BGC and metabolite number.

Ancestral state reconstruction revealed a potential scenario for anthraquinone BGC evolution in the *Teloschistales*, consisting of evolutionarily stable core genes and dynamic lineage-specific accessory genes. Five core genes (EthD, PKSI, MβL, ABC transporter, Methyltransferase) were present in the common ancestor of the *Teloschistales* with secondary losses being rare. In contrast, accessory genes are gained and lost more frequently. Accessory genes contain domains indicative of tailoring functions (i.e. not part of the core machinery for compound skeleton production and transport) and so, as discussed above, may be under less selective pressure to be retained. Apart from the ABC transporter gene, unclustered homologs of all the genes within the cluster are present in the genomes of non-*Teloschistales* lichen mycobionts. This supports the findings of Chapter 2, which suggest that anthraquinone BGCs evolved by shuffling existing genes into novel combinations rather than de novo BGC evolution.

The current dataset and results focus on variation within the *Teloschistales*. A critical next step is to generate paired data for further outgroup taxa to have greater certainty of the evolutionary events that led to the ancestral BGC of the *Teloschistales* crown group. For example, whilst the EthylD, PKSI and MβL are reconstructed as clustered in the common ancestor of all taxa, it is unclear in which order the other core genes were clustered. Particular attention should be paid to sampling lichens within *Dothideomycetes* and *Lichinomycetes* classes, as these are the two most diverse lichen-containing classes not sampled here.

In this dataset, most of the variation in anthraquinone BGCs between species relates to accessory functions, whilst core biosynthetic genes are generally conserved. Therefore, BGC evolution may be more linked to variation in BGC regulation or compound transport than compound structure. Enrichment analysis supports this, as besides a few exceptions, *Teloschistales* lichens, on the whole, do not appear to be enriched in anthraquinone diversity compared to other metabolite groups. Furthermore, modelling showed no difference in the rate of evolution of the number of anthraquinone-like metabolite features per sample between *Teloschistales* and outgroup taxa and no correlation between anthraquinone

evolution and the evolution of individual genes within its corresponding BGC. Therefore the observed variation in anthraquinone BGC gene content is most likely linked to when, where and how anthraquinones are produced rather than which anthraquinones.

The distribution of ABC transporters is of note. Outside the *Teloschistales*, an ABC transporter within an anthraquinone BGC (state clustered in Figure 4.6) is observed in the *Psoraceae* and unclustered in *Pseudevernia consocians* (*Parmeliaceae*). Ancestral state reconstruction suggests the ABC transporter was unclustered in the ancestor of these taxa. This may indicate that the *Psoraceae* independently evolved an anthraquinone BGC with an ABC transporter. *Psoraceae* ABC proteins contain unique protein domains not seen in *Teloschistales*, which further supports this idea. Interestingly, *Psoraceae* lichens often occupy similar ecological niches and geographical distributions to the *Teloschistaceae*, and *Protoblastenia* (*Psoraceae*) species, are commonly misidentified as *Teloschistaceae*. Therefore, adaptations to similar evolutionary pressures may have led to ecological, morphological and genomic convergence.

ABC transporters have also been reported in the lichens *Evernia prunastri* (*Parmeliaceae*) (Pizarro et al. 2020), *Cladonia uncialis*, *C. grayi* (*Cladoniaceae*) (Bertrand et al. 2018b; Armaleo et al. 2019), and *Peltigera membranacea* (*Peltigeraceae*), however not as part of anthraquinone BGCs (i.e. unclustered in regards to my ancestral state reconstruction). As all these taxa are in the *Lecanoromycetidae* subclass, the ABC transporter may have been present in the ancestor of the clade. However, my sampling of the sister subclasses *Umbilicariomycetidae* and *Ostropomycetidae* is limited, and as ancestral state reconstruction is sensitive to missing data (Salisbury et al. 2001), I cannot conclude this definitively.

Within the *Teloschistaceae*, certain lineages like *Pyrenodesmia* appear to have secondarily lost the ability to produce anthraquinones (Frolov et al. 2020). However, my results show that *Pyrenodesmia variabilis* retains the typical anthraquinone BGC. Furthermore, the GNPS feature-based molecular networking identified a feature in the *P. variabilis* metabolome matching the anthraquinone parietin. Therefore the apparent loss of anthraquinones in this lineage may be a downregulation of anthraquinone production, resulting in concentrations below the detection thresholds of traditional chemical spot tests and TLC.

BiG-SCAPE revealed additional BGC families that may be involved in anthraquinone

biosynthesis in non-*Teloschistales* taxa and a handful of *Teloschistales*. Although they may produce similar end products, the vast difference in gene content and structure suggests they evolved independently of the typical anthraquinone BGCs. Given the relatively simple molecular structure of anthraquinones and the fact that alternative anthraquinone pathways have already been described in non-lichenised *Lecanoromycetes* (Jahn et al. 2017), I believe convergent compound evolution is plausible. Furthermore, given that emodin is a common intermediate in many fungal biosynthetic pathways, convergent anthraquinone evolution could evolve through the loss of genes for the final stages of the pathway, leaving the anthraquinone as the end product. The PKSI and MβL genes of these additional BGC families are still homologous to those in the typical *Teloschistales* anthraquinone BGCs, suggesting that although the BGCs may have originated independently, they still utilise some of the same genetic ingredients. A similar scenario was proposed for GAL cluster evolution across the fungal tree of life (Slot et al. 2010).

Paired genome-metabolome datasets are powerful tools for investigating how genomic variation impacts the metabolite phenotype. The results from this chapter suggest a complex interplay between genome and metabolome and reject the idea of a one-to-one relationship between BGCs and metabolites. This highlights the importance of multi-omics approaches, as inferring metabolic capacity purely from genomic data may misrepresent the actual metabolites an organism is producing.

Lichens represent a promising system for exploring further questions in fungal secondary metabolome evolution. As slow-growing organisms, they are unlikely to show major seasonal variability in metabolite production (Grube et al. 2013) whilst still possessing diverse metabolic profiles. Furthemore, their conspicuous perennial thalli make repeated sampling readily attainable. They are also one of the only groups of fungi constantly exposed to the atmosphere (Spribille et al. 2022), which presents an attractive system for exploring how climatic changes may affect the genome and metabolome. Their relatively close relation to model *Ascomycota* like *Aspergillus* also provides a wealth of data and methods to analyse and interpret results. Therefore, I hope this chapter provides inroads towards including lichen-forming fungi within broader studies on the evolution of fungal secondary metabolism.

Chapter 5

Lichens have evolved self-resistance mechanisms to withstand the antifungal effects of anthraquinones

5.1 Abstract

Fungi produce an array of bioactive secondary metabolites. These compounds are often broadly cytotoxic inhibiting universal cellular processes and binding to essential proteins. As a consequence, metabolite-producing fungi have evolved resistance mechanisms to avoid self-toxicity. Lichen-forming fungi are a diverse group known for their rich profiles of bioactive metabolites. Despite this, self-resistance mechanisms in lichens have not been directly investigated. Here we focused on cytotoxic anthraquinone pigments in the *Teloschistaceae* family as a case study to understand self-resistance evolution in lichens. Previous work characterising the anthraquinone biosynthetic gene cluster suggested ABC transporters as potential self-resistance adaptations. As fungi often deploy multiple self-resistance mechanisms, we asked whether *Teloschistaceae* lichens evolved additional adaptations to cope with anthraquinone toxicity. First, we determined the cellular target of anthraquinones. Since anthraquinones are known inhibitors of the DNA repair enzyme topoisomerase II in humans, we asked whether this extends to fungal topoiso-

merase II. Using a yeast DNA topoisomerase II relaxation assay, we show that emodin, a key lichen anthraquinone, inhibits fungal topoisomerase II, whilst its methylated derivative parietin does not. Subsequent LCMS analysis demonstrated that parietin is produced in much higher concentrations than emodin in lichen-forming fungi under both symbiotic and aposymbiotic conditions. Together, these results suggest that *Teloschistaceae* lichens avoid emodin topoisomerase II inhibition through methylation to parietin. Additionally, we observe a methyltransferase within the anthraquinone gene cluster, which supports this hypothesis. Finally, using comparative genomics, we identified unique topoisomerase II mutations under strong positive selection in the *Teloschistaceae*. Protein modelling shows that three mutations fall within the ATPase gate, the domain responsible for enzyme catalytic activity and the predicted binding region for emodin in mammal systems. These observed mutations may confer resistance by either preventing emodin from binding or altering topoisomerase II catalytic activity. Therefore *Teloschistaceae* lichens have achieved anthraquinone self-resistance through the evolution of efflux pumps, compound methylation and resistant target enzymes. This suite of self-resistance mechanisms likely allowed anthraquinones to become the dominant metabolite in the *Teloschistaceae* and a key adaptation for colonising extreme environments.

5.2 Introduction

Fungi rely on producing potent bioactive molecules to defend themselves against competitors, grazers and pathogens. Their potency has long been known and used by humans and has led to some of our most effective medicines (Guzmán et al. 1998; Bills et al. 2016; Prescott et al. 2023). Although some fungal secondary metabolites have highly specific mechanisms, most are non-specific targeting universal cellular processes such as the electron transport chain and DNA repair or disrupting general cellular features like cell membranes (Bills et al. 2016). As a result, they are broadly cytotoxic and metabolite-producing fungi must evolve self-resistance mechanisms to avoid suffering the toxic effects themselves. As mentioned in Chapter 1.1, many of these protective mechanisms are encoded within the biosynthetic gene cluster (BGC) of the toxin itself. These are commonly grouped into three main categories: 1) compound transport, where toxins are shuttled out of the cell, 2) target duplication, where the cellular target is duplicated to absorb or resist

some of the compound intracellularly, and 3) compound detoxification, where toxins are converted into less toxic derivatives (Keller 2015).

Lichens produce a particularly high diversity of bioactive secondary metabolites (Boustie et al. 2005) that are known to be toxic to animals and other microbes. Despite this, beyond the occasional observation of transporters and methyltransferases in lichen BGCs (Dal Grande et al. 2018; Singh et al. 2021b; Singh et al. 2021a; Si et al. 2023; Llewellyn et al. 2023b), self-resistance mechanisms in lichens have not been directly investigated. In this chapter, I aim to investigate these mechanisms for anthraquinones, a key group of lichen metabolites. Anthraquinones are bioactive pigments used in agriculture as fungicides and pesticides and in medicine as anticancer agents (Hildebrandt et al. 2018; Malik et al. 2021). As discussed in Chapter 1.3, lichen anthraquinones are most abundant and conspicuous in the *Teloschistaceae* family. Therefore, the *Teloschistaceae* represents a promising model for exploring anthraquinone self-resistance in fungi. Furthermore, with the new genome-level data presented in previous chapters, we can now investigate the genetic basis of potential self-resistance adaptations.

Due to their prevalence in lichen thalli, emodin and parietin are the most widely studied lichen anthraquinones. Emodin is a common intermediate in many fungal biosynthetic pathways but is rarely seen as an end product, most likely due to its toxicity. Parietin is a derivative of emodin and one of the most common anthraquinones in the *Teloschistaceae*. Although their principal function in lichens is as photoprotective pigments (Gauslaa et al. 2003; Solhaug et al. 2004; Solhaug et al. 2010), their bioactivity suggests they may have additional antimicrobial functions. Given their apparent bioactivity and the fact that *Teloschistaceae* lichens produce them in such high concentrations, it stands to reason that these fungi must have evolved effective self-resistance mechanisms to cope with toxic anthraquinones.

The identification of an ABC transporter within the *Teloschistaceae* anthraquinone BGC (see Chapter 2) presented a first potential self-resistance mechanism, an example of the aforementioned 'compound transport' category. Genes for efflux pump transporters are often observed in fungal BGCs. Expressing transporter genes simultaneously with metabolite production prevents the accumulation of toxic compounds within the cell. ATP Binding Cassette (ABC) and Major Facilitator Superfamily (MFS) transporters are the two most common transporters encoded within fungal BGCs, and both can transport

a wide range of substrates (Keller 2015). This target promiscuity may facilitate BGC evolution as a single ancestral transporter protein gene can evolve to form part of various BGCs producing different metabolites. Furthermore, target promiscuity may allow these transporters to protect against exogenous toxins from other antagonistic fungi (Dubey et al. 2014).

Since fungi often evolve suites of resistance mechanisms, I hypothesise that *Teloschistaceae* lichens evolved additional toxicity avoidance mechanisms besides compound transport. Target duplication, the second category described by Keller (2015), is where toxin-producing fungi evolve to produce resistant duplicates of the enzyme that a toxic compound targets. This can be seen in the fungal BGCs for fumagillin and lovastatin (Wiemann et al. 2013; Nyilasi et al. 2014). This additional copy can contain point mutations that reduce binding affinity to the compound. Furthermore, as the additional copies bind the toxin, they leave other enzymes free to continue cellular processes.

Before we can investigate this second category of self-resistance mechanisms, we must first determine the cellular mode of action of anthraquinone toxicity in fungi, i.e. which fungal enzymes do toxic anthraquinones target if any. In mammalian cells, emodin inhibits the essential DNA repair enzyme topoisomerase II (Mueller et al. 1999b; Li et al. 2010a). As topoisomerase II cleaves and reconnects double-stranded DNA, enzyme inhibition can lead to irreparable DNA damage. Topoisomerase II inhibitors are typically classified as either catalytic inhibitors or poisons, depending on their mode of action. Catalytic inhibitors block topoisomerase II activity completely, whereas poisons trap topoisomerase II and DNA in a tripartite complex. These complexes prevent DNA religation, which causes double-strand DNA breaks (Okoro et al. 2023). Emodin inhibition most likely occurs by blocking ATP hydrolysis, but evidence exists for it also stabilising DNA-Topoisomerase II complexes (Li et al. 2010a). Therefore it shows signatures of both catalytic inhibition and poisoning. Mueller et al. (1999a) suggested that parietin is less toxic than emodin in mammalian cells, but topoisomerase II inhibition was not directly investigated.

To date, there is no direct evidence regarding the cellular toxicity of anthraquinones within a fungal system and whether topoisomerase II inhibition extends to lichen anthraquinones like parietin. Therefore we first performed a topoisomerase II inhibition assay of lichen anthraquinones in a yeast model system to determine their mode of action. Since topoisomerase II enzymes are highly conserved across eukaryotes, I hypothesise that

lichen anthraquinones will also inhibit yeast topoisomerase II enzymes.

After compound transport and resistant target evolution, we aimed to explore compound detoxification, the third main category of self-resistance adaptations. BGCs containing detoxification enzymes, such as oxidoreductases and methyltransferases, are common in fungi. These alter the compound structure in a way that either prevents it from binding to its target or reduces its affinity. This adaptation is seen in the gliotoxin BGC in *Aspergillus fumigatus* (Scharf et al. 2010) and the geodin BGC in *A. nidulans* (Nielsen et al. 2013). A methyltransferase gene was consistently found within the anthraquinone BGCs of *Teloschistaceae* lichens (see Chapter 4, Figure 4.5). As both emodin and its methylated derivative parietin have been reported in *Teloschistaceae* lichens (Søchting 1997), I aimed to test the hypothesis that *Teloschistaceae* lichens avoid toxic emodin buildup via methylation to parietin. If this were the case, I would expect weaker topoisomerase II inhibition for parietin compared to emodin and higher concentrations of parietin than emodin in producing lichens.

Finally, based on the prior hypothesis that anthraquinones target DNA repair, I hypothesised that genes involved in this cellular process in anthraquinone-rich *Teloschistaceae* lichens would be under different evolutionary pressures than those same genes in non-*Teloschistaceae* lichens. Specifically, I asked whether topoisomerase II and other genes related to DNA repair contain unique mutations under positive selection in the *Teloschistaceae* and whether this increases the resistance of *Teloschistaceae* topoisomerase II to inhibitor compounds.

5.3 Methods

Yeast DNA topoisomerase II relaxation assay

The assays described in this section were performed in collaboration with Professor Anthony Maxwell at the John Innes Centre, UK. To investigate whether *Teloschistaceae* lichens have evolved resistant target enzymes, we first determined the cellular mode of action of anthraquinone toxicity. To do this, we assayed lichen compounds using a yeast DNA topoisomerase II system. *Saccharomyces* yeasts belong to the same phylum as *Teloschistaceae* (*Ascomycota*), and as topoisomerase II is a highly conserved protein, we assume protein structure and function to be similar to that of *Teloschistaceae* lichens. I prepared

solutions in DMSO (1mg/mL) of two lichen anthraquinones, parietin and emodin, and one lichen non-anthraquinone, usnic acid, from pure compounds. We then assessed the ability of these compounds to inhibit yeast DNA topoisomerase II using a relaxation assay (Inspiralis Ltd. Norwich, UK). The known topoisomerase inhibitor etoposide was used as a positive control. Yeast DNA topoisomerase II was obtained as a gift from James Berger (Johns Hopkins, USA). Yeast topoisomerase II DNA relaxation assays were carried out using supercoiled plasmid pBR322 and assay buffer (20 mM Tris·HCl (pH 7.9), 200 mM KCl, 10 mM MgCl2, 1.5 mM ATP, 4% (v/v) glycerol), supplied by Inspiralis Ltd. Assays were run at 30°C for 30 mins, and analysed by electrophoresis on 1% agarose gels, stained using ethidium bromide.

Detection & quantification of emodin & parietin

Given my hypothesis that anthraquinone self-resistance occurs through methylation of toxic emodin to less toxic parietin, I expect higher concentrations of parietin than emodin in producing lichens. To test this, I performed compound detection and quantification of emodin and parietin on a paired dataset of symbiotic and aposymbiotic material from 18 samples of lichen-forming fungi (Table A.11). To obtain aposymbiotic material, I produced axenic cultures of the 18 samples using an adapted protocol following the 'spore-shooting' method of Yoshimura et al. (2002) and McDonald et al. (2013); apothecia were excised from clean lichen thalli and fixed to the lid of sterile Petri dishes with either sellotape or Vaseline. The base of the Petri dishes contained 2% malt extract agar medium (MEA). Plates were sealed and left upside down in a flow hood overnight to trigger forcible spore discharge. Plates were checked for spores using a dissecting microscope and if spores were seen, the lid was replaced with a clean one. Plates were left to grow for between six months and one year, removing contaminants or re-plating where necessary.

I aimed to harvest 10 mg of cultured material for metabolite extraction where possible. However, given the variable growth rates of different strains, for those not reaching 10 mg I harvested as much material as possible and adjusted the volume of methanol solvent to achieve 10 mg of wet sample per 0.5 mL of methanol. Sterile 2 mL Eppendorf tubes were weighed twice before and after transferring material to calculate the sample wet weight. For symbiotic material, I separated approximately 10 mg of clean lichen thallus and/or apothecia from the same samples used for axenic cultures. Lichen samples were weighed,

ground and extracted in the same way as the cultured material. The corresponding volume of methanol was added (0.5 mL per 10 mg biomass) and samples were ground using a sterile plastic pestle. Samples were vortexed thoroughly every 10 minutes for one hour and then centrifuged for 2 minutes at 13,000 rpm. The methanol solvent was then transferred to a 0.3 mL insert glass LCMS vial. Extractions were concentrated by fully evaporating the solvent using a Genevac^{TM} Concentrator EZ-2 (SP Industries Inc., USA) for one hour on a medium boiling setting. Once fully evaporated, samples were resuspended in 50 µL of methanol at 4°C overnight.

To calculate concentrations of emodin and parietin, I produced standard dilutions of pure emodin in 100% methanol and pure parietin in 10% methanol 90% DMSO. Solvents were chosen to achieve optimal dissolution. Nine two-fold serial dilutions were prepared ranging from 1 mg/mL to 1.95 µg/mL to produce a standard calibration curve covering the expected concentration range of lichen metabolites.

Lichen and culture extractions and standard dilutions of emodin and parietin were then analysed using liquid-chromatography mass-spectrometry (LCMS). Extractions were run on a VelosPro with the aid of Dr Rui Fang (Royal Botanic Gardens Kew) using the following settings. Blank methanol (HPLC grade) was used as a background control. We performed LCMS analyses on an UltiMate 3000 Standard (SD) HPLC system (Thermo Scientific, USA) coupled to a 100 Hz diode array detector (DAD) and an Ion trap Velos Pro (Thermo Scientific, USA) mass spectrometer. The chromatographic system consists of a Luna C18 column (150 mm x 3 mm i.d., 3µm, Phenomenex, USA) using a mobile phase gradient of 0:90:10 to 90:0:10 [MeOH: water: acetonitrile + 1% formic acid] over 20 minutes. 90% MeOH was held for a further 5 minutes and then returned to initial conditions over 2 minutes, with column temperature at 30°C (flow rate: 400 µL/min). UV detection was recorded between 190 and 450 nm.

We performed mass spectrometry detection in positive and negative ionization modes using the full scan and data-dependent MS2 and MS3 acquisition methods. Total Ion Current (TIC) chromatograms were obtained over the range of 125-2000 m/z using a spray voltage of +3.0 kV and -2.5 kV for the positive and negative ionization modes, respectively. Additional parameters include capillary temperature, 300°C; heater temperature, 300°C; sheath gas flow rate, 60; auxiliary gas flow rate, 20; automatic gain control (AGC) target, 3.0e4 (Full scan) and 1.0e4 (MSn); normalized collision energy for MSn, 35 eV; minimal

signal required, 500 and isolation width, 4. Nitrogen was used as the drying, nebulizer and fragmentation gas.

I detected the presence of emodin and parietin from the LCMS output data using Xcalibur v4.0 software (ThermoFisher Scientific). The standard emodin sample displayed a clear peak at ∼21.5 minutes retention time (RT) and 269-270 m/z in negative ion mode with UV maxima at 440 nm. The parietin standard displayed a peak at ∼22.7 RT and 285-286 m/z in MS+ with UVmax at 437 nm. Two standard calibration curves were plotted for each compound using the MS and UV peak areas and the known concentrations of the serial dilutions. Emodin and parietin presence in the lichen samples were detected by searching for MS-, MS+ and UV peaks that matched the standards. Emodin and parietin concentration, y, was calculated from the peak area, x, using the equation $y = mx + c$, where m is the gradient of the calibration curve and c is the y-intercept of the calibration curve. Results were plotted in R v4.3.0 using the ggplot2 package (Wickham 2016).

To confirm the presence of parietin within living fungal cells, I performed fluorescent microscope imaging of axenic cultures from one anthraquinone-producing lichen species and one non-producing species. I first determined whether parietin autofluorescences by calculating its absorption and emission spectra from a standard 1 mg/mL sample of parietin in DMSO. In a transparent 96-well 100 µL plate (Thermofisher, UK), I prepared six serial dilutions of the parietin standard, diluting by half each step. The plate was read in a Tecan Infinite M200 plate reader (Tecan Trading AG, Switzerland) with 5 nm absorbance scan steps from 400 to 1000 nm. This showed maximum absorption at 435 nm, which was used to determine the max emission in a black fluorescence plate. Max emission was seen at 620 nm. For fluorescence imaging, I mounted approximately one forceps pinch of living mycobiont culture on a glass slide in distilled water. I then examined preparations using a Zeiss Axio Imager 7.2 with Axiocam 506 mono camera (Zeiss, Germany) and images were captured with an oil-immersed 100× objective. I captured autofluorescence images under a 430 nm light filter cube and manually optimized images for best contrast and brightness (ensuring the same settings for both lichens). I exported images as TIFFs.

Selection analysis of stress response genes

Finally, given the hypothesis that anthraquinones affect DNA repair, I tested whether genes involved in this process were under different evolutionary pressures in *Teloschistaceae*

versus non-*Teloschistaceae* lichens. Although topoisomerase II appears to be the most likely target, I also investigated related DNA repair genes to see if the DNA repair process is more widely affected. Along with topoisomerase II, I included the genes for the heat-shock protein Hsp90, which often acts as a regulator of DNA damage response pathways (Pennisi et al. 2015), and RAD52, which repairs DNA breaks through homologous recombination (Nitiss 2009).

I pulled these three genes from a dataset of 117 whole mycobiont genomes. This dataset consisted of the 112 genomes from Chapters 2, 3, and 4 (77 *Teloschistaceae* and 35 non-*Teloschistaceae*) and five newly sequenced non-*Teloschistaceae* lichen genomes covering the families *Ramalinaceae*, *Megasporaceae*, *Parmeliaceae*, *Psoraceae* and *Pyrenulaceae*. For the newly sequenced genomes, I extracted, sequenced and assembled four using short-read Illumina sequence data as in Chapter 2.3 and tested the sequencing of one taxon using long-read Oxford Nanopore technology (ONT). I chose the species *Ramalina farinacea* for ONT due to its large thallus size, as current protocols require ∼400 mg of input material. ONT was not feasible on a broad scale for this project due to the scarcity of material for most taxa studied. Nonetheless, it was undertaken for *R. farinacea* as an example of gold-standard lichen metagenome sequencing to be expanded in future work. See Appendix A.5 for full methods and accompanying published genome note of *Ramalina farinacea* (Llewellyn et al. 2023a).

I predicted mycobiont genes and identified orthologues for Illumina genomes as in Chapter 2.3 (see Supplementary Table A.12 for sample localities). For *Ramalina farinacea* sample information, see Appendix A.5. New WGS data and annotated genome assemblies for Illumina genomes are available on GenBank under the BioProject accession PRJNA991097. Data for *Ramalina farinacea* are available under BioProject accession PRJNA890456.

For the 117 whole mycobiont genomes, I converted genome annotations into a blast database and then used query proteins to search for the three genes of interest using blastp via BLAST+ (Camacho et al. 2009). I chose query protein sequences from published high-quality annotations of *Aspergillus nidulans* FGSC A4 and *Aspergillus fumigatus* Af293. Query NCBI accessions are as follows: Hsp90 (XP_681538.1), RAD52 (XP_752129.1), Topoisomerase II (XP_663010.1). I filtered the results for matches >35% identity and >40% sequence coverage and converted them to fasta files using SeqKit's grep function

(Shen et al. 2016).

I aligned protein sequences using MUSCLE5 (Edgar 2022) and pulled the corresponding unaligned CDSs from the funannotate gene predictions. I used these as input for the command line implementation of PAL2NAL v14 (Suyama et al. 2006) to convert aligned amino acids into aligned CDSs. I manually inspected the alignments for errors.

As selection analysis requires a phylogenetic tree, I reconstructed a phylogenomic species tree for the full 117 taxon dataset as in Chapter 2.3 (filtering for 50% single-copy orthologues, aligning with MAFFT, trimming with TrimAl and concatenated tree reconstruction with IQTree). There is still some debate regarding whether species trees or gene trees are more appropriate for selection analysis (Jeffares et al. 2015; Álvarez-Carretero et al. 2023). In this case, as topoisomerase II, *Hsp90*, and *RAD52* are relatively short, single-copy genes I opted for the species tree as it is likely to be more robust. Furthermore, preliminary gene tree versus species tree comparisons showed very few differences.

Some of the three genes could not be found in all lichen genome annotations. As the taxa in the species tree and protein alignments have to match, I removed any missing taxa from the species tree used for each gene; this was done using a custom R script implementing the keep.tip function of ape v5.6-1 (Paradis et al. 2019). The branch leading to the *Teloschistaceae* was marked with '#1' to signify it as the foreground branch.

For each gene, I calculated the ratio of synonymous (d_N) to non-synonymous (d_S) mutations using CODEML in PAML v.4.9j (Yang 2007). A d_N/d_S (ω)=1 signifies a neutral mutation. A $\omega<1$ signifies a deleterious mutation; thus, purifying selection will reduce its fixation. A $\omega>1$ suggests an advantageous mutation, which will be fixed at a higher rate than a synonymous mutation, and is thus a potential sign of positive selection.

To test for differences in selective pressures between *Teloschistaceae* and non-*Teloschistaceae* genes I used both branch and branch-site models. For the branch model, my null model (M0) assumes that ω does not vary across lineages in the species tree nor sites in the alignment. The alternative model allows for selective pressures to differ between a foreground branch (in my case, the *Teloschistaceae* family) and background branches (the rest of the tree). For the branch-site model, my null model is a modified M2a model where ω of the foreground branch is fixed to 1. For the branch-site alternative model, ω can vary between foreground and background branches and also among sites in the alignment, which allows us to test for positive selection affecting particular sites in the protein specifically

in the *Teloschistaceae* (Yang et al. 2002; Zhang et al. 2005; Yang et al. 2005). I calculated ω and the log-likelihood for each gene under each of the models (null branch model (M0), alternative branch model, null branch-site model (M2a), alternative branch-site model). PAML '.ctl' files for the four models can be found here: `https://github.com/theo-llewellyn/LichenSelfResistance/tree/main/selection_analysis/TEMPLATE`.

To compare the null models to the alternative models I calculated the likelihood ratio test (LRT) statistic using the equation $2\Delta\ell = 2(\ell_1 - \ell_0)$, where ℓ_0 is the log-likelihood for my null models (M0 or M2a) and ℓ_1 is the log-likelihood for the alternative models (alternative branch or alternative branch-site). The LRT statistic was then compared to a χ^2 distribution with the degrees of freedom equalling the difference between the number of parameters of the null and alternative models. For the branch-site model, I then used the Bayes Empirical Bayes (BEB) method within PAML to determine which sites may be under positive selection on the foreground branch.

All bioinformatics scripts for genome assembly, phylogenomics and selection analysis can be found here: `https://github.com/theo-llewellyn/LichenSelfResistance`.

Topoisomerase II protein modelling

The *in silico* protein structure modelling described in this section was run with the aid of Valerij Talagayev and Prof Gerhard Wolber (Freie Universität Berlin). To understand how *Teloschistaceae* topoisomerase II structure compares to that of yeast and how the observed mutations affect topoisomerase function, we performed *in silico* structural modelling of *Teloschistes chrysophthalmus* topoisomerase II. I chose the *Teloschistes chrysophthalmus* strain LIQ125TCHR as the genome assembly was high-quality and displayed all the positively selected site mutations identified via BEB. Also, as the genome was derived from axenic mycobiont culture, there was less risk of contaminant fungal DNA.

I first predicted the main protein domains using InterPro (Blum et al. 2021) via the InterProScan v5.61-93.0 web interface. For *in silico* structure modelling, we then converted the DNA sequence into amino acids using 'The Sequence Manipulation Suite' (Stothard 2000). InterPro results showed the protein consisted of three main domains: the ATPase, DNA-gate and C-Terminal Domain (CTD). The CTD region was excluded from subsequent steps as no reliable templates exist. We generated 3D structures using the Alphafold python package (Jumper et al. 2021), using default settings and setting the max template

option to the first day of structure generation. We first obtained structures for monomers, which were subsequently used to generate dimer structures by applying the multimer option of Alphafold (Evans et al. 2022). The dimer protein model was visualised using PyMol v2.5.5 (DeLano 2002).

5.4 Results

Emodin inhibits fungal DNA topoisomerase II whilst parietin does not

To investigate anthraquinone self-resistance mechanisms, we first determined the cellular target of these compounds. Using a yeast DNA topoisomerase II relaxation assay, we confirm that emodin inhibits fungal topoisomerase II relaxation similarly to etoposide, a known topoisomerase II inhibitor which we used as a positive control (Figure 5.1A). Parietin and the non-anthraquinone usnic acid do not show topoisomerase II inhibition at 120 μM and 100 μM, respectively (Figure 5.1A). A more detailed assay with emodin suggests the IC_{50} (the concentration of compound needed to inhibit enzyme activity by half) is around 5 μM (Figure 5.1B).

Parietin is produced in higher concentrations than emodin

I successfully isolated lichen mycobionts in aposymbiotic axenic culture from 18 samples covering at least eleven species across the families *Caliciaceae*, *Lecanoraceae*, *Ramalinaceae*, *Teloschistaceae*, and *Thelotremataceae* (Table A.11). Some samples could only be identified to genus level and will need barcoding to confirm species epithet. This led to a final paired dataset of 36 samples: 18 aposymbiotic cultures and 18 lichen thalli. Figure 5.2A shows an example of an aposymbiotic culture and lichen thallus for *Xanthoria parietina* (*Teloschistaceae*).

I detected emodin in five cultures and 15 lichen thalli, and parietin in three cultures and 11 thalli (Figure 5.2B). Samples, where compounds were detected, showed either both emodin and parietin or just emodin. Parietin was primarily restricted to the *Teloschistaceae* except for one *Physcia* lichen (LIQ255PHYSP).

In all samples where both compounds were detected, parietin was found in higher concentrations than emodin (Figure 5.2C). This was the same for both aposymbiotic cultures and lichen thalli. Emodin peak areas were mostly below the quantification threshold for

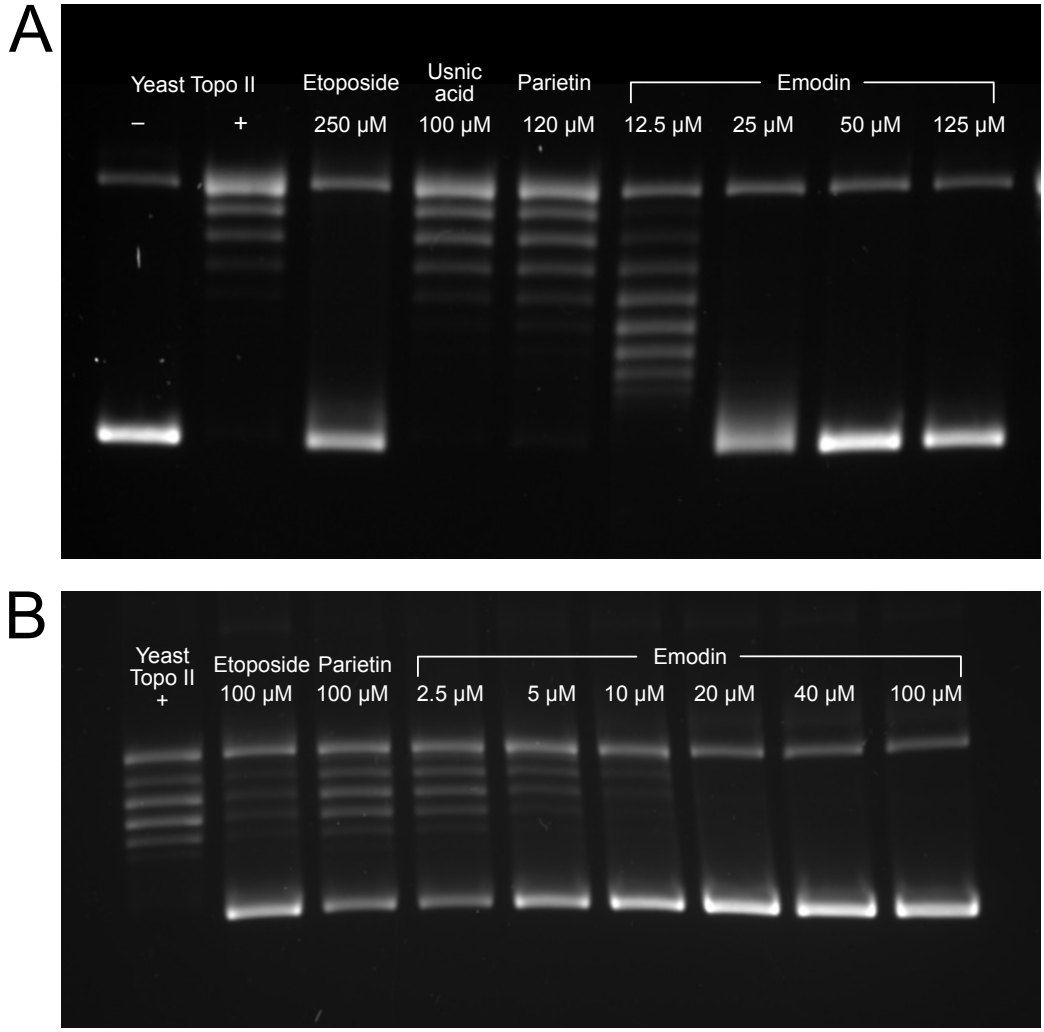


Figure 5.1: Gel electrophoresis photos showing results of relaxation assays of lichen compounds with yeast DNA topoisomerase II system. **A**) Full assay testing lichen compounds usnic acid, parietin and emodin. The first two columns are controls with supercoiled plasmid pBR322 but without yeast topoisomerase II (-) and with both supercoiled plasmid pBR322 and yeast topoisomerase II but no compounds (+). Etoposide is a positive control known topoisomerase II inhibitor. **B**) Detailed assay of emodin across more concentrations showing IC_{50} around 5 μM emodin

the standard curve (shown as zero values in Figure 5.2C), and concentrations could only be directly calculated for five lichen thalli. Parietin was present in much higher concentrations and could be quantified for three cultures and 11 lichen thalli. As most emodin concentrations could only be recorded as 'below limit of quantification', statistical tests could not be performed. There was some discrepancy between the concentrations calculated from the MS and UV standard curves. I opted for the concentrations calculated from the MS peaks, as they had both higher detection rates and coefficients of determination for the standard curves (parietin:$R^2 = 0.9909$, emodin:$R^2 = 0.9656$).

The highest emodin concentration was found in the crustose *Caloplaca* sp. 1 lichen thallus sample (LIQ110CASP) at 1.6304 µg/mg fungal biomass ($mean = 0.0865\mu g/mg \pm$

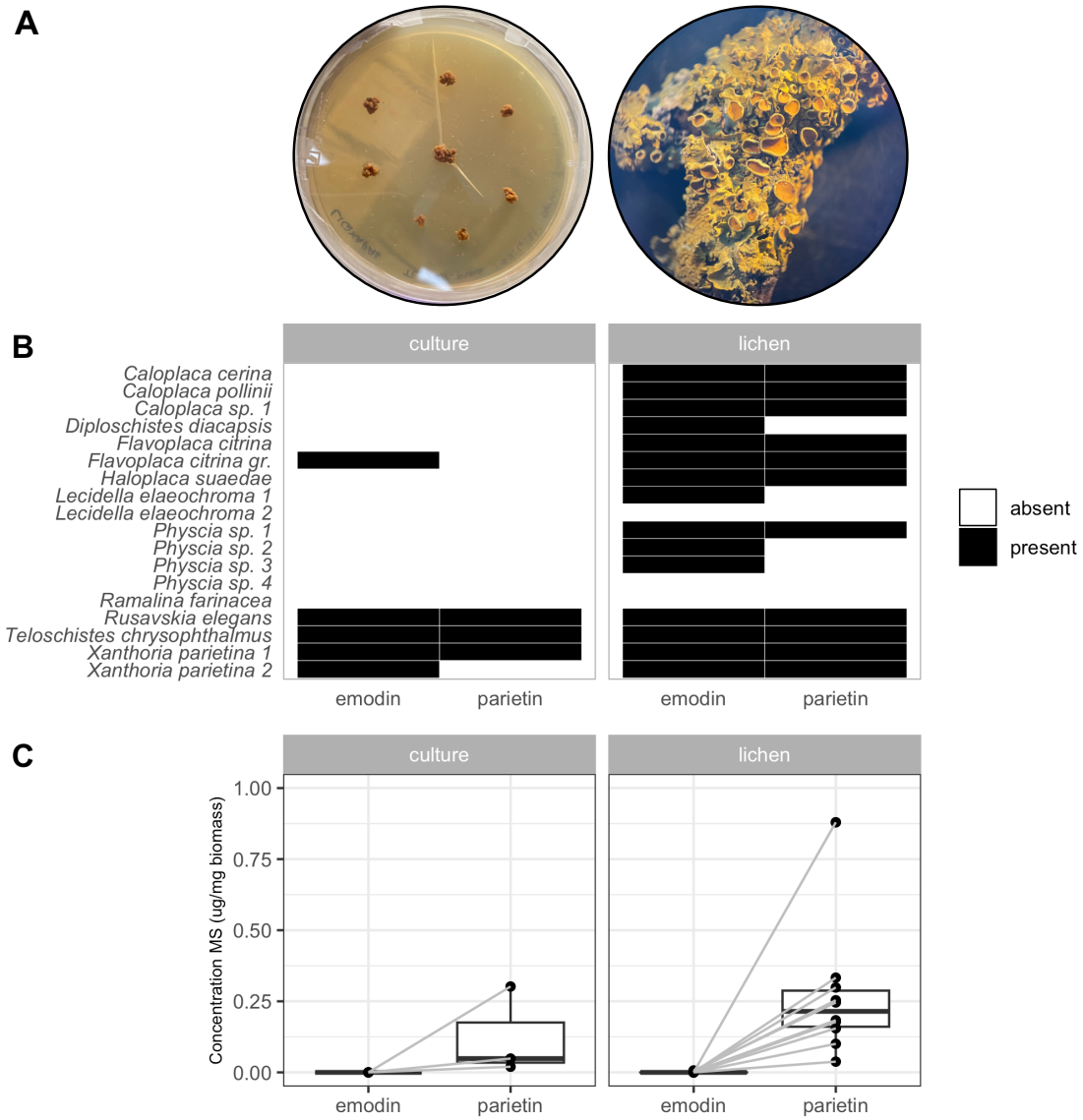


Figure 5.2: Presence and concentration of emodin and parietin in cultured aposymbiotic lichen mycobionts and whole lichen thalli using Liquid-Chromatography Mass-Spectrometry (LCMS). **A**) Photos showing aposymbiotic culture and lichen thallus of *Xanthoria parietina*. **B**) Presence/absence of emodin and parietin in cultures (left) and lichen thalli (right). **C**) Concentrations of emodin and parietin per mg of culture (left) and thallus (right). Concentrations were calculated from standard curves of pure compound serial dilutions. Parietin quantified from MS+ peak at ∼21.5 RT and 285 m/z, emodin quantified from MS- at ∼22.7 RT and 269 m/z. Concentrations below limit of quantification are shown as 0 values. For readability, right-hand plot (lichen) does not show outlier *Caloplaca* sp. 1 as parietin concentration was significantly higher than other samples at $2.38 \mu g/mg$

0.3739). The highest parietin concentration was found in the same sample at 2.3847 µg/mg fungal biomass ($mean = 0.3872 \mu g/mg \pm 0.6128$). Due to the low recovery of both compounds in culture, only three comparisons between culture and lichen material of the same species could be made. Parietin concentration, as determined by MS+ 285 peaks, was ∼45x higher in lichen thallus compared to culture for *Xanthoria parietina* and ∼4x higher for *Teloschistes chrysophthalmus* (Supplementary Figure A.15). For *Rusavskia elegans* the opposite was seen with the culture showing ∼2x higher parietin concentration than the

lichen thallus (Supplementary Figure A.15).

Fluorescence imaging of a representative parietin-producing and non-producing aposymbiotic culture showed that parietin produces a strong autofluorescence signal under 430 nm light (Figure 5.3). For the producing sample *Flavoplaca citrina* (accession LIQ2785) autofluorescence could be seen throughout the hyphae (Figure 5.3A). For the non-producing sample *Lecanora chlarotera* (*Lecanoraceae*), almost no autofluorescence was detected (Figure 5.3B).

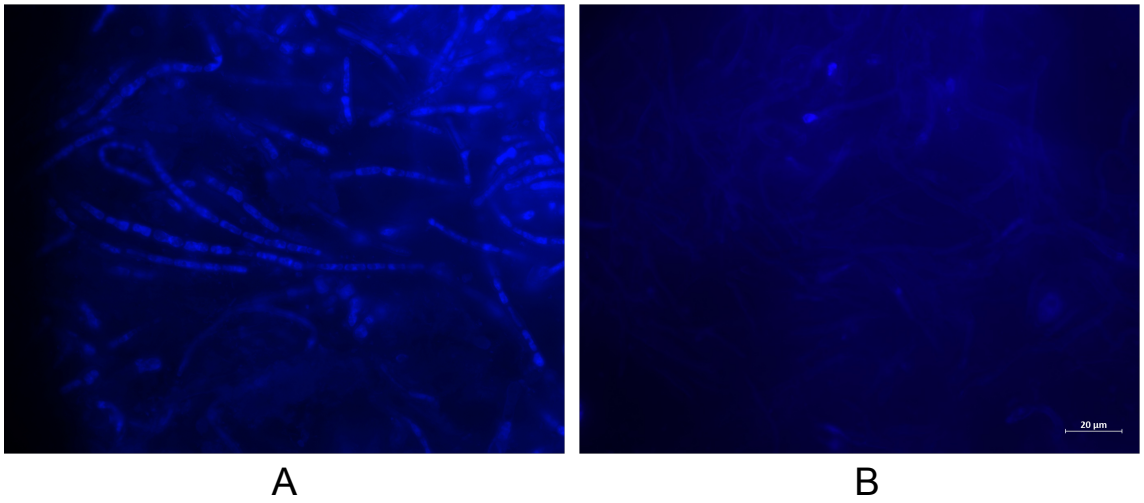


Figure 5.3: Fluorescence microscope images of lichen mycobiont cultures under 430 nm wavelength light showing autofluorescence of parietin within fungal hyphae. **A**) *Flavoplaca citrina* sample accession LIQ2785. **B**) *Lecanora chlarotera*

DNA repair genes are under purifying negative selection as a whole but contain sites under strong positive selection in the *Teloschistaceae*

I undertook selection analysis to identify signatures of selection in DNA repair genes in a dataset of 117 lichen-forming fungal genomes. Five novel genomes were generated for this chapter, including a long-read reference-quality genome for *Ramalina farinacea*. All novel genomes were highly complete and contiguous. For assembly quality metrics for novel short-read Illumina genomes, see Supplementary Table A.13. For *Ramalina farinacea*, see Appendix A.5.

From the 117 lichen-forming fungal genomes, I extracted the annotated protein sequences for Hsp90, RAD52 and topoisomerase II. For the three genes, CODEML selection analysis using the branch model and *Teloschistaceae* as the foreground branch showed all estimated ω values were less than one. This suggests negative purifying selection for the

gene as a whole on both foreground and background branches. For Hsp90, ω was higher for the *Teloschistaceae* branch (ω=0.06470) compared to the rest of the tree (ω=0.02650). For RAD52, ω was lower for the *Teloschistaceae* branch (0.06202) compared to the rest of the tree (0.21422). For topoisomerase II, ω was lower for the *Teloschistaceae* branch (0.05946) compared to the rest of the tree (0.09427). For all three genes, despite results suggesting negative purifying selection on both the *Teloschistaceae* branch and background branches, the LRT showed that foreground and background ω values were still significantly different (see M0 vs branch in Table 5.1). Therefore, the branch model was a better fit than the null M0 model.

Table 5.1: Likelihood ratio tests (LRT) of branch and branch-site models. ℓ_0 = log-likelihood of null model, ℓ_1 = log-likelihood of alternative model. $2\Delta\ell$ = Likelihood Ratio Test statistic ($2 \times (\ell_1 - \ell_0)$).

Gene	**Model comparison**	**log-likelihood (ℓ)**	$2\Delta\ell$	χ^2 **p-value**
Hsp90	M0 vs branch	$\ell_1 = -48516.7325$ $\ell_0 = -48518.9985$	4.5319	0.0333*
	M2a vs branch-site	$\ell_1 = -48262.6453$ $\ell_0 = -48274.5280$	23.7654	<0.0001*
RAD52	M0 vs branch	$\ell_1 = -92308.3866$ $\ell_0 = -92318.4285$	20.0839	<0.0001*
	M2a vs branch-site	$\ell_1 = -89641.0827$ $\ell_0 = -89641.0827$	0	1
Topoisomerase II	M0 vs branch	$\ell_1 = -174649.1613$ $\ell_0 = -174651.7684$	5.2141	0.0224*
	M2a vs branch-site	$\ell_1 = -173045.1741$ $\ell_0 = -173051.8271$	13.3056	0.0003*

Note: df = 1, critical values $\chi^2_{1,5\%} = 3.84$ and $\chi^2_{2,5\%} = 5.99$. * denotes significant p-value

For the branch-site model where ω can vary both between the foreground/background branches and between sites in the alignment, CODEML estimates ω for four site classes; site class 0 contains sites under purifying selection in both the *Teloschistaceae* branch and background branches, site class 1 contains sites under neutral evolution, site class 2a contains sites under positive selection on the *Teloschistaceae* branch and purifying selection on background branches, and site 2b contains sites under positive selection on the *Teloschistaceae* branch and neutral evolution on background branches.

Proportions of sites within each site class and ω values for those sites on both the *Teloschistaceae* branch and background branches are shown in Table 5.2. For Hsp90 ∼96% of sites were under purifying selection in all branches (site class 0), ∼3% were undergoing neutral evolution (site class 1), and ∼1% were under positive selection for the *Teloschistaceae* lineage but not in background lineages (site classes 2a and 2b). For

positively selected sites in the *Teloschistaceae* lineage ω=88.79355. For RAD52, ∼65% of sites were under purifying selection, and ∼35% were under neutral evolution. There was no difference in foreground and background ω estimates. For topoisomerase II, ∼83% of sites were under purifying selection, ∼16% were under neutral evolution, and ∼1.5% were under positive selection for the *Teloschistaceae* branch but not in background branches. For positively selected sites in the *Teloschistaceae* lineage ω=8.84082. The LRT comparing the branch-site model results to the null model (M2a) where $\omega_2 = 1$ is fixed, showed that the branch-site model was a significantly better fit for Hsp90 and topoisomerase II but not for RAD52 (see M2a vs branch-site in Table 5.1).

Table 5.2: CODEML branch-site model results showing proportion of amino acid sites in each site class and corresponding dN/dS (ω) values. Site class 0 = sites under purifying selection in all branches, Site class 1 = sites undergoing neutral evolution, Site class 2a = sites under positive selection for the *Teloschistaceae* but purifying selection for outgroups, Site class 2b = sites under positive selection for *Teloschistaceae* but undergoing neutral evolution in outgroups.

Gene	Site class	Proportion	Background ω	Foreground ω
Hsp90	0	0.95919	0.02261	0.02261
	1	0.03118	1.00000	1.00000
	2a	0.00933	0.02261	88.79355
	2b	0.00030	1.00000	88.79355
RAD52	0	0.64550	0.09468	0.09468
	1	0.35450	1	1
	2a	0	0.09468	1
	2b	0	1	1
Topoisomerase II	0	0.82766	0.07667	0.07667
	1	0.15682	1	1
	2a	0.01305	0.07667	8.84082
	2b	0.00247	1	8.84082

For Hsp90 and topoisomerase II, BEB method calculated posterior probabilities for positively selected sites in the *Teloschistaceae* branch. For Hsp90, of the six identified positive sites, four had significant posterior probabilities (P>95%). For topoisomerase II, of the 16 identified positive sites, five had significant posterior probabilities (Table 5.3).

Since topoisomerase II had a higher number of positively selected sites and is now a confirmed target for anthraquinones in fungi, we used protein modelling to explore whether the topoisomerase II mutations at the positively selected sites shown in Table 5.3 may confer resistance to emodin toxicity. Protein domain annotation with InterProScan of *Teloschistes chrysophthalmus* topoisomerase II showed the two canonical regions known from yeast topoisomerase II (Schmidt et al. 2012) consisting of four domains: the ATPase

Table 5.3: Bayes Empirical Bayes (BEB) analysis of sites under positive selection in the *Teloschistaceae* foreground lineage for the DNA repair genes encoding Hsp90 and topoisomerase II.

Gene	Site	Posterior probability
Hsp90	8	0.995
	10	0.975
	476	0.998
	641	0.994
Topoisomerase II	477	0.982
	532	0.967
	595	0.952
	1058	0.958
	1904	0.979

gate, which consists of an ATPase domain belonging to the GHKL family and a transducer domain, and the DNA gate, which consists of the TOPRIM and CAP domains. The five positively selected sites fall within the GHKL domain, the transducer domain, the CAP region and the disordered CTD region (Figure 5.4A). Figure 5.4B shows the aligned amino acids at the first four positively selected sites. I excluded the fifth site as it fell within the CTD region, which is highly variable and thus harder to align confidently. The 3D structure of the CTD has also yet to be resolved, so I could not place the mutated site within a protein model. The first site (position 330 in unaligned amino acid sequence for LIQ125TCHR: position 477 in the alignment) shows all *Teloschistaceae* having a Tyrosine (Y), an aromatic side chain amino acid, whilst background lineages have amino acids with positively charged side chains (Lysine (K) or Arginine (R)), except for *Diploschistes diacapsis*, which also has a Tyrosine. For the second site (385:532), *Teloschistaceae* mostly have amino acids with non-polar aliphatic side chains (Valine (V) or Isoleucine (I)). In contrast, background lineages mostly have amino acids with positively charged side chains (R and K). For the third (444:595) and fourth sites (901:1058), *Teloschistaceae* have amino acids with either polar uncharged or non-polar aliphatic side chains, whereas background lineages have predominantly polar uncharged amino acids.

To investigate *Teloschistaceae* topoisomerase II structure and how the observed mutations affect topoisomerase function, we performed *in silico* structural modelling of the *Teloschistes chrysophthalmus* topoisomerase II homodimer. The modelled structure shows the two key regions, the ATPase and DNA gates (Figure 5.4C). Based on the similarity to the yeast topoisomerase II model of Schmidt et al. (2012), the 3D structure seems plausible. The model also shows the positions of the observed mutations within 3D protein

space. Mutation TYR-330 sits on the boundary of the GHKL and transducer domains within the ATPase gate in close proximity to the other monomer. Mutations ILE-385 and SER-444 are clustered in the transducer domain, with ILE-385 on the external surface of the protein and SER-444 on the internal surface. Mutation SER-901 is relatively isolated from the others on the internal surface of the CAP domain.

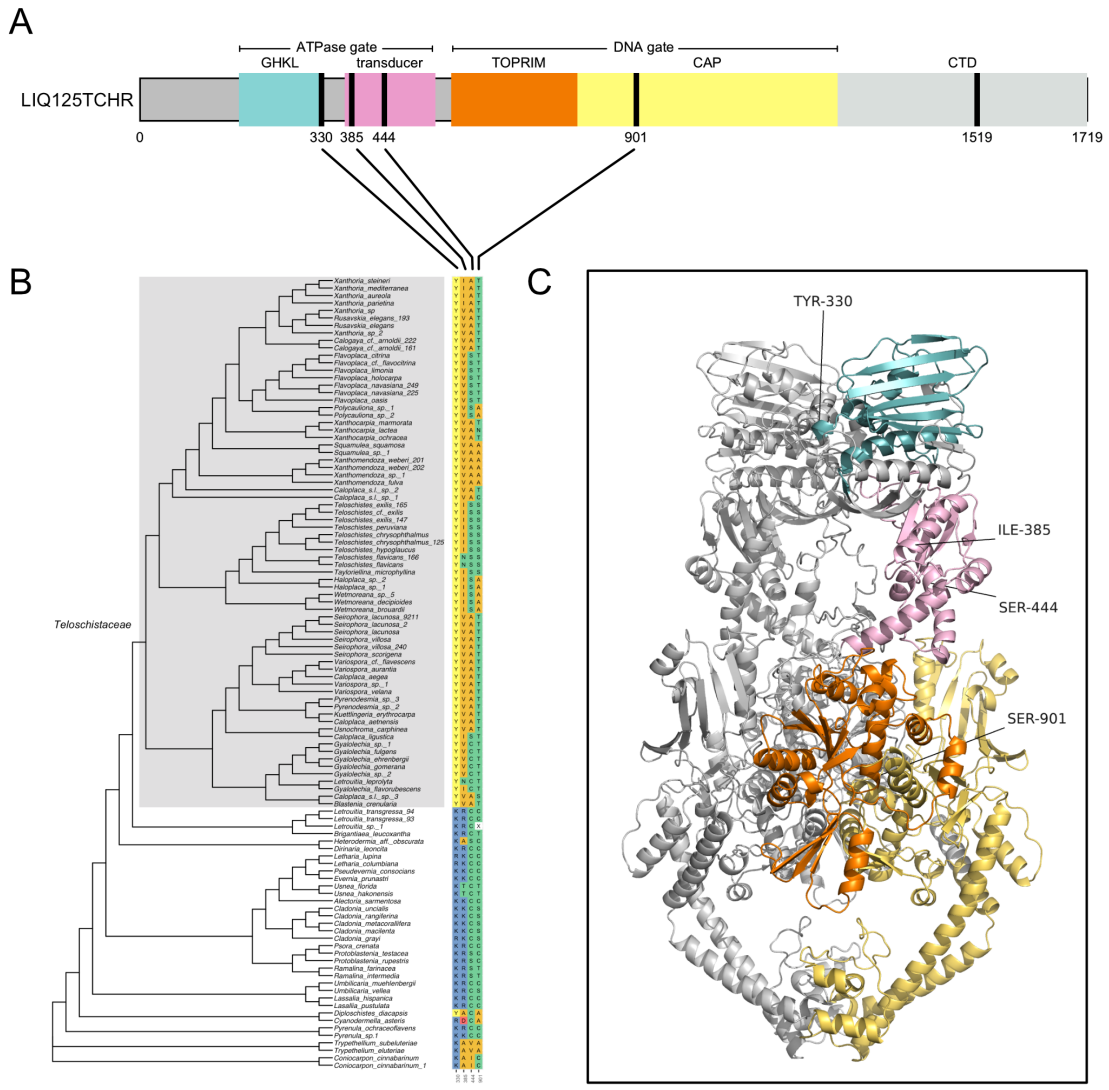


Figure 5.4: Topoisomerase II structure showing sites under positive selection in the *Teloschistaceae* lineage. **A**) Length and arrangement of domains/regions as annotated from InterProScan. Sites identified by Bayes empirical Bayes (BEB) method within PAML to be under positive selection in the *Teloschistaceae* are marked with horizontal black lines. Numbers show their corresponding amino acid position in the unaligned *Teloschistes chrysophthalmus* sequence (accession LIQ125TCHR). **B**) Whole-genome phylogenetic tree with foreground *Teloschistaceae* lineage highlighted. Columns show 4/5 positively selected sites with their corresponding amino acids. Amino acids are coloured according to side chain chemistry: yellow = aromatic, blue = +ve charged, orange = non-polar aliphatic, green = polar uncharged, red = -ve charged. **C**) *In silico* model of *T. chrysophthalmus* topoisomerase II homodimer, with mutations labelled. Protein domains of one monomer are coloured as in subplot **A**, the other monomer is coloured grey. Colours in **A** and **C** do not correspond to those in **B**.

5.5 Discussion

Lichen-forming fungi produce a wide range of bioactive molecules at high concentrations, many of which are toxic to other organisms, including other fungi. Given the toxicity-avoidance hypothesis (McGary et al. 2013; Rokas et al. 2018), lichen mycobionts will likely have evolved adaptations to avoid damaging themselves. Prior to this study, however, self-resistance mechanisms had not been thoroughly investigated in lichens. Here, we investigated self-resistance evolution in response to anthraquinone production in the lichen family *Teloschistaceae*. The first main category of self-resistance adaptations, compound transport, was already identified in previous chapters by the detection and description of ABC transporters. Therefore, we aimed to investigate the two other main categories: detoxification and resistant target enzymes.

To explore detoxification adaptations, we first determined the enzyme target of toxic anthraquinones. Using a yeast DNA topoisomerase II relaxation assay, we show that emodin inhibits fungal topoisomerase II, whilst parietin does not. Additionally, we observe much higher concentrations of parietin than emodin in *Teloschistaceae* lichens under both symbiotic and aposymbiotic conditions. Therefore, we propose detoxification via methylation of emodin to parietin as a key self-resistance adaptation in *Teloschistaceae* lichens that produce high concentrations of anthraquinones. A methyltransferase gene within *Teloschistaceae* anthraquinone BGCs provides the genetic basis for this.

Following detoxification, we explored the third category of self-resistance adaptations, resistant target enzyme evolution. We observe unique mutations in the target enzyme topoisomerase II that are under positive selection in the *Teloschistaceae*. Protein modelling shows three mutations occur within the ATPase gate that may confer emodin resistance. This differs slightly from the target duplication strategy mentioned in the introduction as the resistant topoisomerase II is not encoded within the anthraquinone BGC itself. Therefore *Teloschistaceae* lichens appear to have achieved anthraquinone resistance through adaptations covering all three categories of self-resistance: the evolution of ABC transporter efflux pumps, compound detoxification via methylation, and resistant topoisomerase II enzymes.

Given how structurally similar emodin and parietin are, it is surprising how differently they affect topoisomerase II. Understanding this difference requires a more thorough

investigation of how anthraquinones bind to fungal topoisomerase II. Li et al. (2010a) predicted that in humans emodin binds to topoisomerase II by forming polar hydroxyl bonds between its positions 1,3 hydroxy groups and the topoisomerase ATPase domain. Additionally, in general mutagenicity assays comparing different anthraquinones, only anthraquinones with two hydroxy groups at positions 1,3 showed toxicity (Tikkanen et al. 1983; Westendorf et al. 1990). Therefore, methylation to parietin, which replaces the position 3 hydroxy group of emodin (R-OH) with a methoxy group ($R\text{-}O\text{-}CH_3$), may prevent or weaken bonds between parietin and the ATPase gate thus explaining the observed lack of topoisomerase II inhibition.

One limitation of this study was using a yeast model system to determine the target of anthraquinones rather than investigating it directly using topoisomerase II from lichen-forming fungi. Whilst necessary due to the difficulties of genetic manipulation with lichens, it is possible that emodin inhibition in lichens differs from inhibition in yeast. However, considering how conserved topoisomerase II is across eukaryotes, and the similarity of our *in silico* topoisomerase II model (Figure 5.4C) to that of yeast, I believe the observed topoisomerase II inhibition also applies to lichen-forming fungi. The selection analysis further supports topoisomerase II as a target, as anthraquinone-rich *Teloschistaceae* lichens show unique positively selected mutations, that may confer resistance. Topoisomerase II is an essential protein in fungi (Holm et al. 1985). Therefore continuous exposure to topoisomerase II inhibitors like emodin will likely create strong selective pressure for resistance. This may explain the high ω values seen for positively selected sites in the *Teloschistaceae*.

The observed mutations cluster within the ATPase gate, which controls catalytic activity. Whilst the ATPase gate is the predicted binding region for emodin (Li et al. 2010a), we must determine the exact binding site within the ATPase gate to understand if and how these mutations confer resistance. Topoisomerase II mutations conferring resistance to poisons are much better understood than mutations conferring resistance to catalytic inhibitors like emodin. Poison-resistance mutations can occur in both the ATPase and DNA gates. ATPase gate mutations generally confer resistance by reducing topoisomerase catalytic activity (Nitiss 2009; Wu et al. 2011) whilst DNA gate mutations confer resistance by preventing the formation of enzyme-DNA-poison tripartite complexes (Nitiss et al. 1994; Sabourin et al. 1998).

For catalytic inhibitors, whilst cell lines with resistance mutations have been described (Hasinoff et al. 1997; Wessel et al. 1999; Jensen et al. 2000; Wessel et al. 2002), the mechanisms by which mutations confer resistance are unknown. Some studies suggest these mutations confer resistance by preventing inhibitors from binding, while others suggest mutations causing reduced or even increased topoisomerase activity is the mechanism (see Beck et al. (1999) and references within). As none of the mutations from these studies match those seen in *Teloschistaceae* topoisomerase II, further work is needed to understand how resistance may occur in our case. As a starting point, we are currently running *in silico* structure generation with and without mutations 330, 385 and 444 to understand their effects on potential binding sites and enzyme catalytic activity. To investigate this more directly, future research should focus on engineering yeast strains with the observed mutations and measuring their emodin binding affinity and catalytic activity. This would show whether the observed mutations confer resistance by preventing emodin binding or changing enzyme turnover.

Some resistance mechanisms act as potentiators for further resistance adaptations. For example, in a controlled evolution experiment using *Staphylococcus aureus*, the efflux pump *norA* increased the benefit of resistant DNA topoisomerase mutations meaning that resistance mutations were more likely to go to fixation, thereby providing even stronger antibiotic resistance (Papkou et al. 2020). A similar pattern was seen with transcriptional regulators in *Pseudomonas* bacteria (Gifford et al. 2018). Therefore, two key questions arise: Which of the observed mechanisms arose first? Did earlier mechanisms potentiate the later ones? The ancestral state reconstruction in Chapter 4 showed that an anthraquinone BGC containing both an ABC transporter and a methyltransferase was likely to be present in the ancestor the *Teloschistales* order. The observed topoisomerase II mutations are shared only within the *Teloschistaceae* family, suggesting they were fixed after the divergence of *Teloschistaceae* from its sister clade *Letrouitineae*. Therefore, ABC transporters and methyltransferases may have potentiated the evolution of resistant topoisomerase II enzymes in the *Teloschistaceae* by making the observed resistance mutations more likely to reach fixation.

Although *Letrouitinieae* species can produce anthraquinones, they are restricted to reproductive structures rather than throughout the whole lichen, as seen in the *Teloschistaceae*. Consequently, there may have been less selective pressure to evolve resistant

topoisomerase II enzymes than in the *Teloschistaceae*. What is not clear is whether prior topoisomerase II mutations had already conferred a degree of anthraquinone resistance before the *Letrouitineae-Teloschistineae* split. To understand this, one could search for additional positively selected mutations shared across the whole *Teloschistales* and remodel the topoisomerase II structure with these changes to see whether they could confer resistance.

In this study, I focused on resistance mechanisms within the BGC and target enzymes. However, additional self-resistance mechanisms can be found outside the BGC. For example, confining toxin production to cellular compartments is common in fungi (Keller 2015). These compartments can be simple intra and extracellular lipids droplets (Chang et al. 2015) or specialised organelles like peroxisomes and toxisomes (Keller 2015). Fungi often combine cellular compartmentalisation with the BGC-related resistance mechanisms into a sophisticated armoury of toxicity-avoidance adaptations. With the fluorescent microscopy approach used here, I could not determine the localisation of anthraquinone biosynthesis at a subcellular level. Mass Spectrometry imaging methods such as laser desorption/ionization time-of-flight mass spectrometry (LDI-MS) have shown great potential for spatial mapping of lichen metabolites (Le Pogam et al. 2016) and may allow us to test this hypothesis further.

This chapter investigated self-resistance mechanisms in the mycobiont. Since lichens are stable microbial associations, we may expect self-resistance in the photobiont and microbiome symbionts too. This has been observed in some lichen-associated bacteria that can transform the lichen metabolite usnic acid into less toxic derivatives (Noël et al. 2021). Our use of metagenomics provides the data for investigating this in future work. Lichenicolous fungi specialised to parasite *Teloschistaceae*, such as the *Tremella caloplacae* group, would be a key target for anthraquinone resistance as they actively grow on and degrade *Teloschistaceae* thalli. These antagonistic fungal parasites may have even contributed to the evolution of *Teloschistaceae* anthraquinones in the first place, given the clear antifungal properties of anthraquinones, and the strong selective pressure that a specialist lichen parasite would exert on its host to evolve defence strategies (see Discussion in Chapter 4).

DNA topoisomerases are universal across the tree of life. Therefore, results from this study may have applications beyond lichenology and evolutionary biology. Topoisomerase

II poisons such as etoposide and doxorubicin are crucial first-line cancer drugs. However, given their severe side effects, current research is exploring catalytic inhibitors as alternatives (Okoro et al. 2023). Given the diversity of anthraquinones within the *Teloschistaceae* (Santesson 1970; Søchting 1997; Søchting 2001; Søchting et al. 2002; Arup et al. 2013), these fungi represent a promising source of novel topoisomerase catalytic inhibitors. Additionally, some resistance mechanisms characterised here (efflux pumps, mutations in the drug target) are often seen in drug-resistant human cancer cells (Cox et al. 2016). Therefore, our results could inform future work into how cells evolve resistance to novel topoisomerase II inhibitors. This is pertinent given the frequent emergence of resistance mutations to widely used chemotherapeutic agents like doxorubicin (Cox et al. 2016).

Understanding self-resistance mechanisms in lichens can provide insight into antimicrobial resistance in fungi, more generally. Antifungal resistance currently poses one of the greatest challenges for treating plant and human fungal pathogens (Hof 2001; Perlin et al. 2017; World Health Organization (WHO) 2022). As with drug-resistant cancer cells, the adaptations we observe here are similar to those fungal pathogens use to evade host defences and medical interventions (Hof 2001). Therefore understanding anthraquinone self-resistance in *Teloschistaceae* could help us to predict and combat antifungal resistance adaptations in clinical and agricultural settings. Our results are especially relevant for understanding resistance in crop pathogens like mildews, which are treated with antifungals containing parietin and emodin (Ma et al. 2010; Hildebrandt et al. 2018). Furthermore, by understanding these mechanisms in non-model fungi such as lichens, we may be better prepared for future emerging fungal pathogens, which may arise from under-studied clades and could thus respond differently to current antifungal treatments.

Chapter 6

Summary & Conclusions

In this thesis, I explored two main questions: how do adaptive metabolic traits arise in lichen-forming fungi, and what are the evolutionary processes that shape metabolite diversification? Focusing on the case study of anthraquinone evolution in the *Teloschistales*, I aimed to address these questions by pursuing four main objectives:

Objective 1 - Characterise the biosynthetic genes responsible for anthraquinone production in the *Teloschistales*

Objective 2 - Produce a robust genome-scale time tree for the *Teloschistales* to resolve relationships between higher taxa and provide baseline data for phylogenetic comparative analysis

Objective 3 - Understand the evolutionary processes that shape metabolite diversification and determine how genomic changes affect the chemical phenotype

Objective 4 - Investigate self-resistance adaptations in the *Teloschistaceae* as a consequence of toxic anthraquinone production

Anthraquinone evolution in the *Teloschistales* is a pertinent case study to explore my overarching research questions for multiple reasons. Firstly, the established association between *Teloschistales* anthraquinones and arid ecosystem colonisation (Gaya et al. 2015) provided an initial hypothesis that anthraquinone production is a key adaptive trait in this clade. Whilst this is not the sole reason for their evolutionary success in these habitats, the ubiquity of anthraquinones in the group suggests they played and continue to play an important role. Secondly, despite being one of the most diverse lichen lineages, only three genomes had been published as of October 2019. Therefore, the *Teloschistales* represented a major gap in the *Lecanoromycetes* class for genome-scale data. Thirdly, as

the *Teloschistales* has a prevalence of small, crustose species—often the most challenging to culture—they presented a good opportunity to explore the potential of culture-independent approaches like metagenomics. This is relevant because a significant proportion of lichen diversity comprises crustose microlichens. In addition, the stability and relatively simple chemical structure of anthraquinones made them an attractive case study, as they were easier to recover with standard metabolite extractions. Finally, the fact that anthraquinones are common intermediates in many fungal biosynthetic pathways greatly facilitated the characterisation and annotation of biosynthetic gene clusters and metabolites, as I could incorporate highly curated data from model fungi.

In Chapter 2, I developed a novel lichen-specific metagenomics pipeline, which I used to assemble 24 new, high-quality lichen mycobiont genomes. Introduced by Tyson et al. (2004), microbial genome assembly from metagenomic reads has rapidly become standard procedure for studying the genomes of complex bacterial and archaeal communities. Though only recently applied to lichens, this approach is revolutionising lichen genomics. Even within the space of this project, the number of metagenome-assembled lichen mycobiont genomes (MAGs) has surpassed that of culture-derived genomes (Greshake Tzovaras et al. 2020; McKenzie et al. 2020; Tagirdzhanova et al. 2021; Singh et al. 2021b; Singh et al. 2022; Resl et al. 2022). Bypassing the time and labour-intensive culturing step allows for a significantly higher throughput of samples meaning that ambitious sequencing projects like the Vertebrate Genome Project (Rhie et al. 2021) and the 5000 arthropod genomes initiative (i5K) (i5K Consortium 2013) may soon be within reach for lichens.

However, lichens are still an underrepresented group in fungal genomics. A concerted community effort to sequence lichen metagenomes could help to address this. Leaning towards this bioinformatic approach comes with its own challenges, a key one being the lack of fully automated metagenomics pipelines for eukaryotes, as detailed in Chapter 2.5. Additionally, as lichen metagenomics develops, it is crucial to establish a set of standard criteria for assessing lichen MAG quality. These already exist for bacteria and archaea (Bowers et al. 2017) and could be adapted for lichens, being mindful of the non-trivial differences between prokaryote and eukaryote genomes. Combining standard metrics with ground truth data such as cytometric genome size measurements will ensure the quality and comparability of lichen MAGs, as demonstrated in Appendix A.5 for the lichen *Ramalina farinacea*.

Nevertheless, culture-derived lichen genomics should not be abandoned as it allows us to investigate important questions that metagenomics cannot. Potentially the most

interesting of these relate to the genes and gene products involved in establishing and maintaining symbiosis between lichen partners (Joneson et al. 2011; Armaleo et al. 2019). The first stages of lichen symbiotic establishment have yet to be observed in nature. Therefore, co-culture experiments currently provide the only method for understanding this process.

In Chapter 2, I annotated secondary metabolite biosynthetic gene clusters (BGCs) in a combined dataset of my newly sequenced genomes and publicly available *Lecanoromycetes* genomes. This showed that lichen BGCs are numerous and highly dissimilar. Comparative genomics allowed me to identify putative anthraquinones BGCs in *Teloschistales* genomes and characterise their gene content and structure (**Objective 1**). These results showed that putative anthraquinone BGCs in the *Teloschistales* contain a minimum of four genes, including a unique ABC transporter gene. Core PKS genes were widely conserved across taxa beyond the *Teloschistales*, suggesting anthraquinone BGCs arose by re-shuffling existing core enzyme genes—used by many other clades across the fungal tree of life—with novel accessory genes. This contrasts with results for other lichen metabolites (Pizarro et al. 2020), which point to whole BGC gains or losses as the key genetic process. This suggests that, as with non-lichen-forming fungi, different genetic processes shape the diversification of different metabolite groups. The consistent recovery of an ABC transporter in *Teloschistales* anthraquinone BGCs and the general lack of it in non-*Teloschistales* lichens also provided the first hypothesis for toxicity avoidance (**Objective 4**); ABC transporters would give *Teloschistales* taxa an advantage over closely related lichens by conferring protection against their own anthraquinones.

Until the function of these BGCs can be directly confirmed by heterologous expression, I recommend the use of 'putative anthraquinone BGCs' to describe them. Given the observed similarity to anthraquinone BGCs in non-lichen-forming fungi, my results provide a promising set of targets for such analyses. As mentioned in Chapter 2, genetic manipulation in lichens is still in its infancy, with only three reported cases of successful heterologous expression (Harvey et al. 2018; Kealey et al. 2021; Kim et al. 2021). Though many technical problems still need resolving before this approach can be widely applied to lichens (reviewed in Gill et al. (2023)), heterologous expression has the potential to significantly improve our understanding of the true biosynthetic potential of lichen-forming fungi.

Having demonstrated the efficacy of my metagenomics pipeline, in Chapter 3 I then expanded my sampling to sequence all major lineages in the *Teloschistales*. Here, I re-

ported 59 new lichen mycobiont genomes, which I combined with genomes from Chapter 2 and marker locus data for over 500 taxa to produce a genome-level phylogenetic review of *Teloschistales* systematics (**Objective 2**). My results showed that most genera described since Arup et al. (2013) are not supported and suggest a major taxonomic revision of higher taxa is needed for the clade. Given the taxonomic breadth and robustness of the dataset, I also revisited divergence time estimation using a sequential Bayesian dating approach. I showed that genome-scale data significantly reduces the confidence intervals of node age estimations. Going forward, this will be crucial for interpreting diversification rate analyses as it will allow any observed rate shifts to be more precisely linked to geological events.

To tackle **Objective 3**, I generated novel untargeted metabolome data, which I analysed alongside the genomic dataset and time-calibrated phylogeny produced in Chapters 2 and 3. Trait evolution analysis revealed a complex interplay between genomic and metabolic variation. When summarising metabolome diversity into a dissimilarity matrix between species, no correlation was seen between metabolic variation and BGC dissimilarity or phylogeny. However, the presence-absence patterns of individual metabolites and the overall number of metabolites per species did correlate with phylogenetic relatedness. Ancestral state reconstruction of anthraquinone evolution improved BGC characterisation, revealing stable core genes and dynamic lineage-specific accessory genes (**Objective 1**). Results also strengthened the hypothesis from Chapter 2 that constituent genes of anthraquinone BGCs evolved before the *Teloschistales* diverged and that re-shuffling existing genes led to metabolic innovation in this clade. Together, this indicates that metabolite diversification is shaped by a combination of shared common ancestry and lineage-specific BGC variation. Joint modelling of anthraquinone BGCs and metabolite features suggested that genomic changes have not significantly increased anthraquinone structural diversity within the *Teloschistales* and most likely are linked to regulation and transport. Therefore, the results of Chapter 4 suggest that whilst *Teloschistales* lichens clearly produce a high diversity of anthraquinones, their metabolomes are not more enriched in anthraquinones compared to metabolomes of other anthraquinone-producing lichen clades such as the *Psoraceae* and *Pyrenulaceae*. This reframes the informal notion that *Teloschistales* are the most anthraquinone-rich lichens and suggests that their uniqueness instead lies in the abundance and ubiquity of the anthraquinones they accumulate.

The description of ABC transporters in Chapter 2 prompted a more in-depth investigation of self-resistance mechanisms in *Teloschistales* fungi (**Objective 4**). In Chapter 5,

we confirmed topoisomerase II—an essential DNA repair enzyme—as a cellular target of anthraquinones in fungal systems, and that methylated anthraquinones are less toxic than their unmethylated counterparts. Tests for positive selection in topoisomerase II genes revealed unique sites under strong positive selection in the *Teloschistaceae* family. *In silico* protein modelling suggested these mutations may confer resistance to anthraquinone toxicity through either preventing emodin from binding or altering topoisomerase II catalytic activity. The results in Chapter 5 therefore present the following hypothesised scenario of self-resistance in *Teloschistaceae* lichens: as emodin is produced, it is methylated into a less toxic parietin compound by a methyltransferase encoded within the anthraquinone BGC. The toxic effect of any remaining intracellular emodin is buffered by having resistant topoisomerase II enzymes, which results in reduced DNA damage. Finally, ABC transporter efflux pumps—also encoded within the BGC—transport the anthraquinones out of the cell, where they crystallise into a stable, extracellular layer on the lichen thallus cortex.

Current literature only discusses *Teloschistales* anthraquinones as UV protective compounds (Gauslaa et al. 2003; Solhaug et al. 2004; Solhaug et al. 2010; Gaya et al. 2015). Whilst they are clearly photoprotective, results in Chapter 5 strengthen the evidence for them having an additional antimicrobial function. Anthraquinones show toxicity levels akin to potent topoisomerase II inhibitors like etoposide, and *Teloschistales* fungi have had to evolve a suite of self-resistance adaptations to cope with them. As mentioned in Chapter 1, adaptive traits often evolve in response to changing biotic interactions, leading to coevolution. As lichen-forming fungi grow extremely slowly, they are unable to outgrow antagonists. Therefore, toxic anthraquinones could be an important adaptive trait for deterring competitors and pathogens. Since the mycobiont relies on photosynthates from the photobiont, preventing competitors from overgrowing and blocking light is essential. Furthermore, as lichen fungal pathogens (lichenicolous fungi) often cause extensive host thallus necrosis, we can assume that they create a strong selective pressure for defensive trait evolution. The variation in anthraquinone chemosyndromes across the *Teloschistaceae* could, therefore, reflect adaptation to different antagonistic biotic interactions. Whether anthraquinones are also involved in interactions with mutualistic lichen symbionts remains to be investigated.

In this thesis, I have focused on mycobiont adaptive traits. However, we know there is much more to a lichen than the mycobiont. From the 97 lichen metagenomes presented here, over 500 additional bacterial symbiont genomes were recovered (data not presented,

but the analytical pipeline to assemble, annotate and identify bacterial genomes was developed and can be found at `https://github.com/theo-llewellyn/Lichen_microbiome_analysis`). Given this immense diversity, analysing the biosynthetic potential of these bacteria was beyond the scope of this thesis. Nonetheless, the clear difference between secondary metabolite phenotypes of whole lichen thalli and axenic mycobiont cultures seen in this study and others (Brunauer et al. 2007; Díaz et al. 2020) highlights the importance of studying the entire microbial community within lichens to fully understand secondary metabolism evolution. For example, do the functions of metabolites produced by different symbiont partners complement each other? If so, could this accelerate adaptation by reducing redundant adaptive trait evolution? Additionally, while anthraquinone BGCs appear to have originated in the fungal partner, some potential cases exist of horizontally transferred metabolite traits between lichen mycobionts and bacteria (Schmitt et al. 2009; Parrot et al. 2016). Therefore, to what extent do lichen symbionts share metabolite traits, and can multiple partners contribute to the same secondary metabolic pathways? Finally, given the potent bioactivity of mycobiont metabolites, have these other partners also evolved resistance mechanisms? If that is the case, are obligate members of the symbiosis more resistant than commensal ones?

Having overcome initial difficulties, lichen genomics is now providing a much richer and more complete account of the role of secondary metabolism in fungal evolution and adaptation. The drastic increase in available lichen-forming fungal genomes within the last five years also means there is now no excuse to exclude lichens from wider comparative studies. Furthermore, new lichen genomes present unique opportunities for testing important hypotheses about fungal secondary metabolism, particularly regarding the role of secondary metabolites in establishing obligate mutualisms and adapting to extreme environments. Challenges remain, such as automating metagenomic pipelines, improving genetic manipulation of lichen biosynthetic genes and assessing the impact of non-fungal symbionts. However, if these challenges are addressed, lichen genomics has the potential to leave a lasting impact on the field of fungal evolutionary biology.

References

Agarwal, S. K. et al. (2000). Antifungal Activity of Anthraquinone Derivatives from Rheum Emodi. *Journal of Ethnopharmacology* 72(1-2):43–46. DOI: 10.1016/S0378-8741(00)00195-1.

Allen, J. L., Jones, S. J. M., and McMullin, R. T. (2021). Draft Genome Sequence of the Lichenized Fungus Bacidia Gigantensis. *Microbiology Resource Announcements* 10(44):e00686–21. DOI: 10.1128/MRA.00686-21.

Alneberg, J. et al. (2014). Binning Metagenomic Contigs by Coverage and Composition. *Nature Methods* 11(11):1144–1146. DOI: 10.1038/nmeth.3103.

Álvarez-Carretero, S., Kapli, P., and Yang, Z. (2023). Beginner's Guide on the Use of PAML to Detect Positive Selection. *Molecular Biology and Evolution* 40(4):msad041. DOI: 10.1093/molbev/msad041.

Álvarez-Carretero, S. et al. (2022). A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. *Nature* 602(7896):263–267. DOI: 10.1038/s41586-021-04341-1.

Ament-Velásquez, S. L. et al. (2021). The Plot Thickens: Haploid and Triploid-Like Thalli, Hybridization, and Biased Mating Type Ratios in Letharia. *Frontiers in Fungal Biology* 2. DOI: 10.3389/ffunb.2021.656386.

Amnuaykanjanasin, A. and Daub, M. E. (2009). The Abc Transporter Atr1 Is Necessary for Efflux of the Toxin Cercosporin in the Fungus Cercospora Nicotianae. *Fungal Genetics and Biology* 46(2):146–158. DOI: 10.1016/j.fgb.2008.11.007.

Andrade, A. C. et al. (2000). The Abc Transporter Atrb from Aspergillus Nidulans Mediates Resistance to All Major Classes of Fungicides and Some Natural Toxic Compounds. *Microbiology* 146(8):1987–1997. DOI: 10.1099/00221287-146-8-1987.

Andrews, S. et al. (2015). *Fastqc. a Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics*. Babraham Bioinformatics. URL: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Armaleo, D., Sun, X., and Culberson, C. (2011). Insights from the First Putative Biosynthetic Gene Cluster for a Lichen Depside and Depsidone. *Mycologia* 103(4):741–754. DOI: 10.3852/10-335.

Armaleo, D. et al. (2019). The Lichen Symbiosis Re-Viewed Through the Genomes of Cladonia Grayi and Its Algal Partner Asterochloris Glomerata. *BMC Genomics* 20(1):605. DOI: 10.1186/s12864-019-5629-x.

Arup, U., Søchting, U., and Frödén, P. (2013). A New Taxonomy of the Family Teloschistaceae. *Nordic Journal of Botany* 31(1):016–083. DOI: 10.1111/j.1756-1051.2013.00062.x.

Avise, J. C. and Johns, G. C. (1999). Proposal for a Standardized Temporal Scheme of Biological Classification for Extant Species. *Proceedings of the National Academy of Sciences* 96(13):7358–7363. DOI: 10.1073/pnas.96.13.7358.

Awakawa, T. et al. (2009). Physically Discrete β-Lactamase-Type Thioesterase Catalyzes Product Release in Atrochrysone Synthesis by Iterative Type I Polyketide Synthase. *Chemistry & Biology* 16(6):613–623. DOI: 10.1016/j.chembiol.2009.04.004.

Azzalini, A. (2019). *The R Package 'Sn': The Skew-Normal and Related Distributions Such as the Skew-t.* URL: http://azzalini.stat.unipd.it/SN/.

Barba-Montoya, J., Tao, Q., and Kumar, S. (2021). Assessing Rapid Relaxed-Clock Methods for Phylogenomic Dating. *Genome Biology and Evolution* 13(11):evab251. DOI: 10.1093/gbe/evab251.

Baselga, A. (2010). Partitioning the Turnover and Nestedness Components of Beta Diversity. *Global Ecology and Biogeography* 19(1):134–143. DOI: 10.1111/j.1466-8238.2009.00490.x.

Bateman, A. (2019). Uniprot: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Research* 47(D1):D506–D515. DOI: 10.1093/nar/gky1049. pmid: 30395287.

Battistuzzi, F. U. et al. (2018). RelTime Relaxes the Strict Molecular Clock throughout the Phylogeny. *Genome Biology and Evolution* 10(6):1631–1636. DOI: 10.1093/gbe/evy118.

Beaulieu, J. M., O'Meara, B. C., and Donoghue, M. J. (2013). Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. *Systematic Biology* 62(5):725–737. DOI: 10.1093/sysbio/syt034.

Beaulieu, J. M. et al. (2012). Modeling Stabilizing Selection: Expanding the Ornstein–Uhlenbeck Model of Adaptive Evolution. *Evolution* 66(8):2369–2383. DOI: 10.1111/j.1558-5646.2012.01619.x.

Beavan, A. J. S. et al. (2020). Performance of A Priori and A Posteriori Calibration Strategies in Divergence Time Estimation. *Genome Biology and Evolution* 12(7):1087–1098. DOI: 10.1093/gbe/evaa105.

Beck, W. T. et al. (1999). Tumor Cell Resistance to Dna Topoisomerase Ii Inhibitors: New Developments. *Drug Resistance Updates* 2(6):382–389. DOI: 10.1054/drup.1999.0110.

Beimforde, C. et al. (2014). Estimating the Phanerozoic History of the Ascomycota Lineages: Combining Fossil and Molecular Data. *Molecular Phylogenetics and Evolution* 78:386–398. DOI: 10.1016/j.ympev.2014.04.024.

Bengtsson-Palme, J. et al. (2013). Improved Software Detection and Extraction of Its1 and Its2 from Ribosomal Its Sequences of Fungi and Other Eukaryotes for Analysis of Environmental Sequencing Data. *Methods in Ecology and Evolution* 4(10):914–919. DOI: 10.1111/2041-210X.12073.

Bertrand, R. L., Abdel-Hameed, M., and Sorensen, J. L. (2018a). Lichen Biosynthetic Gene Clusters Part II: Homology Mapping Suggests a Functional Diversity. *Journal of Natural Products* 81(4):732–748. DOI: 10.1021/acs.jnatprod.7b00770.

— (2018b). Lichen Biosynthetic Gene Clusters. Part I. Genome Sequencing Reveals a Rich Biosynthetic Potential. *Journal of Natural Products* 81(4):723–731. DOI: 10.1021/acs.jnatprod.7b00769.

Bertrand, R. L. and Sorensen, J. L. (2018c). A Comprehensive Catalogue of Polyketide Synthase Gene Clusters in Lichenizing Fungi. *Journal of Industrial Microbiology and Biotechnology* 45(12):1067–1081. DOI: 10.1007/s10295-018-2080-y.

Bertrand, Y., Pleijel, F., and Rouse, G. (2006). Taxonomic Surrogacy in Biodiversity Assessments, and the Meaning of Linnaean Ranks. *Systematics and Biodiversity* 4(2):149–159. DOI: 10.1017/S1477200005001908.

Bills, G. F. and Gloer, J. B. (2016). Biologically Active Secondary Metabolites from the Fungi. *Microbiology Spectrum* 4(6):10.1128/microbiolspec.funk-0009–2016. DOI: 10.1128/microbiolspec.funk-0009-2016.

Blanco, O. et al. (2006). Major Clades of Parmelioid Lichens (Parmeliaceae, Ascomycota) and the Evolution of Their Morphological and Chemical Diversity. *Molecular Phylogenetics and Evolution* 39(1):52–69. DOI: 10.1016/j.ympev.2005.12.015.

Blin, K. et al. (2019). Antismash 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Research* 47(W1):W81–W87. DOI: 10.1093/nar/gkz310. pmid: 31032519.

Blomberg, S. P., Garland JR., T., and Ives, A. R. (2003). Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* 57(4):717–745. DOI: 10.1111/j.0014-3820.2003.tb00285.x.

Blum, M. et al. (2021). The Interpro Protein Families and Domains Database: 20 Years On. *Nucleic Acids Research* 49(D1):D344–D354. DOI: 10.1093/nar/gkaa977.

Boddy, L. and Hiscox, J. (2017). "Fungal Ecology: Principles and Mechanisms of Colonization and Competition by Saprotrophic Fungi". In: *The Fungal Kingdom*. Ed. by J. Heitman et al. Washington DC: American Society of Microbiology:pp. 293–308. ISBN: 978-1-55581-958-3. DOI: 10.1128/9781555819583.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30(15):2114–2120. DOI: 10.1093/bioinformatics/btu170. pmid: 24695404.

Borowiec, M. L. (2016). Amas: A Fast Tool for Alignment Manipulation and Computing of Summary Statistics. *PeerJ* 2016(1). DOI: 10.7717/peerj.1660.

Boustie, J. and Grube, M. (2005). Lichens—a Promising Source of Bioactive Secondary Metabolites. *Plant Genetic Resources* 3(2):273–287. DOI: 10.1079/pgr200572.

Bowers, R. M. et al. (2017). Minimum Information About a Single Amplified Genome (Misag) and a Metagenome-Assembled Genome (Mimag) of Bacteria and Archaea. *Nature Biotechnology* 35(8):725–731. DOI: 10.1038/nbt.3893.

Boyce, G. R. et al. (2019). Psychoactive Plant- and Mushroom-Associated Alkaloids from Two Behavior Modifying Cicada Pathogens. *Fungal Ecology* 41:147–164. DOI: 10.1016/j.funeco.2019.06.002.

Boyko, J. D., O'Meara, B. C., and Beaulieu, J. M. (2023). A Novel Method for Jointly Modeling the Evolution of Discrete and Continuous Traits. *Evolution* 77(3):836–851. DOI: 10.1093/evolut/qpad002.

Bräse, S. et al. (2009). Chemistry and Biology of Mycotoxins and Related Fungal Metabolites. *Chemical Reviews* 109(9):3903–3990. DOI: 10.1021/cr050001f.

Brunauer, G. et al. (2007). Alterations in Secondary Metabolism of Aposymbiotically Grown Mycobionts of Xanthoria Elegans and Cultured Resynthesis Stages. *Plant Physiology and Biochemistry* 45(2):146–151. DOI: 10.1016/j.plaphy.2007.01.004. pmid: 17344057.

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and Sensitive Protein Alignment Using DIAMOND. *Nature Methods* 12(1):59–60. DOI: 10.1038/nmeth.3176. pmid: 25402007.

Bungartz, F., Søchting, U., and Arup, U. (2020). Teloschistaceae (Lichenized Ascomycota) from the Galapagos Islands: A Phylogenetic Revision Based on Morphological, Anatomical, Chemical, and Molecular Data. *Plant and Fungal Systematics* 65(2):515–576. DOI: 10.35535/pfsyst-2020-0030.

Bushnell, B. (2014). *BBTools http://bbtools.jgi.doe.gov/*. URL: http://bbtools.jgi.doe.gov/.

Butler, M. A. and King, A. A. (2004). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist* 164(6):683–695. DOI: 10.1086/426002.

Caesar, L. K. et al. (2023). Correlative Metabologenomics of 110 Fungi Reveals Metabolite–Gene Cluster Pairs. *Nature Chemical Biology*:1–9. DOI: 10.1038/s41589-023-01276-8.

Calchera, A. et al. (2019). Biosynthetic Gene Content of the 'Perfume Lichens' Evernia Prunastri and Pseudevernia Furfuracea. *Molecules* 24(1). DOI: 10.3390/molecules24010203.

Calcott, M. J. et al. (2018). Secondary Metabolism in the Lichen Symbiosis. *Chemical Society Reviews* 47(5):1730–1760. DOI: 10.1039/c7cs00431a.

Camacho, C. et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics* 10(1):421. DOI: 10.1186/1471-2105-10-421.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* 25(15):1972–1973. DOI: 10.1093/bioinformatics/btp348. pmid: 19505945.

Chain, E. et al. (1940). Penicillin as a Chemotherapeutic Agent. *The Lancet*. DOI: 10.1016/S0140-6736(01)08728-1.

Chang, W. et al. (2015). Trapping Toxins within Lipid Droplets Is a Resistance Mechanism in Fungi. *Scientific Reports* 5(1) (1):15133. DOI: 10.1038/srep15133.

Chang, Y. et al. (2019). Phylogenomics of Endogonaceae and Evolution of Mycorrhizas within Mucoromycota. *New Phytologist* 222(1):511–525. DOI: 10.1111/nph.15613.

Chesters, D. (2020). The Phylogeny of Insects in the Data-Driven Era. *Systematic Entomology* 45(3):540–551. DOI: 10.1111/syen.12414.

Chevrette, M. G. et al. (2020). Evolutionary Dynamics of Natural Product Biosynthesis in Bacteria. *Natural Product Reports* 37(4):566–599. DOI: 10.1039/C9NP00048H.

Chiang, Y.-M. et al. (2010). Characterization of the Aspergillus Nidulans Monodictyphenone Gene Cluster. *Applied and Environmental Microbiology* 76(7):2067–2074. DOI: 10.1128/AEM.02187-09.

Clauzade, G. and Roux, C. (1985). Likenoj de Okcidenta Eŭropo : Ilustrita Determinlibro. *Bulletin de la Société botanique du Centre-Ouest . Nouvelle série. Numero spécial* 7:1–893.

Costa, F. P., Schrago, C. G., and Mello, B. (2022). Assessing the Relative Performance of Fast Molecular Dating Methods for Phylogenomic Data. *BMC Genomics* 23(1):798. DOI: 10.1186/s12864-022-09030-5.

Cox, J. and Weinman, S. (2016). Mechanisms of Doxorubicin Resistance in Hepatocellular Carcinoma. *Hepatic Oncology* 3(1):57–59. DOI: 10.2217/hep.15.41. pmid: 26998221.

Crawford, J. M. et al. (2009). Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* 461(7267):1139–1143. DOI: 10.1038/nature08475.

Crawford, J. M. and Townsend, C. A. (2010). New Insights into the Formation of Fungal Aromatic Polyketides. *Nature Reviews Microbiology* 8(12):879–889. DOI: 10.1038/nrmicro2465. pmid: 21079635.

Crespo, A. et al. (2010). Phylogenetic Generic Classification of Parmelioid Lichens (Parmeliaceae, Ascomycota) Based on Molecular, Morphological and Chemical Evidence. *Taxon* 59(6):1735–1753. DOI: 10.1002/tax.596008.

Cubero, O. F. and Crespo, A. (2002). "Isolation of Nucleic Acids From Lichens". In: *Protocols in Lichenology*:pp. 381–391. DOI: 10.1007/978-3-642-56359-1_23.

Culberson, C. F. and Culberson, W. L. (1976). Chemosyndromic Variation in Lichens. *Systematic Botany* 1(4):325. DOI: 10.2307/2418700.

Dal Grande, F. et al. (2018). The Draft Genome of the Lichen-Forming Fungus Lasallia Hispanica (Frey) Sancho & A. Crespo. *Lichenologist* 50(3):329–340. DOI: 10.1017/S002428291800021X.

Darriba, D. et al. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution* 37(1):291–294. DOI: 10.1093/molbev/msz189.

Dayrat, B. (2005). Towards Integrative Taxonomy. *Biological Journal of the Linnean Society* 85(3):407–417. DOI: 10.1111/j.1095-8312.2005.00503.x.

Degnan, J. H. and Rosenberg, N. A. (2009). Gene Tree Discordance, Phylogenetic Inference and the Multispecies Coalescent. *Trends in Ecology & Evolution* 24(6):332–340. DOI: 10.1016/j.tree.2009.01.009.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. Version 2.5.5. DeLano Scientific. URL: https://pymol.org/2/.

De Queiroz, K. (2006). The PhyloCode and the Distinction between Taxonomy and Nomenclature. *Systematic Biology* 55(1):160–162. DOI: 10.1080/10635150500431221.

Díaz, E. M. et al. (2020). Axenic Culture and Biosynthesis of Secondary Compounds in Lichen Symbiotic Fungi, the Parmeliaceae. *Symbiosis* 82(1):79–93. DOI: 10.1007/s13199-020-00719-3.

Díaz-Escandón, D. et al. (2022). Genome-Level Analyses Resolve an Ancient Lineage of Symbiotic Ascomycetes. *Current Biology* 32(23):5209–5218.e5. DOI: 10.1016/j.cub.2022.11.014.

Divakar, P. K. et al. (2015). Evolution of Complex Symbiotic Relationships in a Morphologically Derived Family of Lichen-Forming Fungi. *New Phytologist* 208(4):1217–1226. DOI: 10.1111/nph.13553.

Donoghue, M. J. and Gauthier, J. A. (2004). Implementing the PhyloCode. *Trends in Ecology & Evolution* 19(6):281–282. DOI: 10.1016/j.tree.2004.04.004. pmid: 16701272.

Dos Reis, M. and Yang, Z. (2011). Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times. *Molecular Biology and Evolution* 28(7):2161–2172. DOI: 10.1093/molbev/msr045.

Dos Reis, M., Zhu, T., and Yang, Z. (2014). The Impact of the Rate Prior on Bayesian Estimation of Divergence Times with Multiple Loci. *Systematic Biology* 63(4):555–565. DOI: 10.1093/sysbio/syu020.

Dos Reis, M., Donoghue, P. C. J., and Yang, Z. (2016). Bayesian Molecular Clock Dating of Species Divergences in the Genomics Era. *Nature Reviews Genetics* 17(2) (2):71–80. DOI: 10.1038/nrg.2015.8.

Dos Reis, M. et al. (2012). Phylogenomic Datasets Provide Both Precision and Accuracy in Estimating the Timescale of Placental Mammal Phylogeny. *Proceedings of the Royal Society B: Biological Sciences* 279(1742):3491–3500. DOI: 10.1098/rspb.2012.0683.

Dos Reis, M. et al. (2015). Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Current Biology* 25(22):2939–2950. DOI: 10.1016/j.cub.2015.09.066.

Duan, M. et al. (2023). Genome, Transcriptome, and Metabolome Analyses Provide New Insights into the Resource Development in an Edible Fungus Dictyophora Indusiata. *Frontiers in Microbiology* 14. DOI: 10.3389/fmicb.2023.1137159.

Dubey, M. K., Jensen, D. F., and Karlsson, M. (2014). An ATP-Binding Cassette Pleiotropic Drug Transporter Protein Is Required for Xenobiotic Tolerance and Antagonism in the Fungal Biocontrol Agent Clonostachys Rosea. *Molecular Plant-Microbe Interactions®* 27(7):725–732. DOI: 10.1094/MPMI-12-13-0365-R.

Duncan, K. R. et al. (2015). Molecular Networking and Pattern-Based Genome Mining Improves Discovery of Biosynthetic Gene Clusters and Their Products from Salinispora Species. *Chemistry & Biology* 22(4):460–471. DOI: 10.1016/j.chembiol.2015.03.010.

Edgar, R. C. (2022). *High-Accuracy Alignment Ensembles Enable Unbiased Assessments of Sequence Homology and Phylogeny.* DOI: 10.1101/2021.06.20.449169. preprint.

Egbert, S. et al. (2022). Unraveling Usnic Acid: A Comparison of Biosynthetic Gene Clusters between Two Reindeer Lichen (Cladonia Rangiferina and C. Uncialis). *Fungal Biology* 126(11):697–706. DOI: 10.1016/j.funbio.2022.08.007.

El Meouche, I. and Dunlop, M. J. (2018). Heterogeneity in Efflux Pump Expression Predisposes Antibiotic-Resistant Cells to Mutation. *Science (New York, N.Y.)* 362(6415):686–690. DOI: 10.1126/science.aar7981.

Elix, J. A. and Stocker-Wörgötter, E. (2008). "Biochemistry and Secondary Metabolites." In: *Lichen Biology*. Cambridge: Cambridge University Press:pp. 104–133. ISBN: 978-0-521-87162-4 978-0-521-69216-8.

Elix, J. A. (2014). A Catalogue of Standardized Chromatographic Data and Biosynthetic Relationships for Lichen Substances. *Third Edition. Published by the author, Canberra (Australia)* (September):1–323.

Elmore, M. H. et al. (2015). Clustering of Two Genes Putatively Involved in Cyanate Detoxification Evolved Recently and Independently in Multiple Fungal Lineages. *Genome Biology and Evolution* 7(3):789–800. DOI: 10.1093/gbe/evv025.

Emms, D. M. and Kelly, S. (2019). OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biology* 20(1). DOI: 10.1186/s13059-019-1832-y. pmid: 31727128.

Eskola, M. et al. (2020). Worldwide Contamination of Food-Crops with Mycotoxins: Validity of the Widely Cited 'FAO Estimate' of 25%. *Critical Reviews in Food Science and Nutrition* 60(16):2773–2789. DOI: 10.1080/10408398.2019.1658570. pmid: 31478403.

Evans, R. et al. (2022). *Protein Complex Prediction with AlphaFold-Multimer*. DOI: 10.1101/2021.10.04.463034. preprint.

Ezard, T., Fujisawa, T., and Barraclough, T. (2013). *R Package Splits: SPecies' LImits by Threshold Statistics, Version 1.0-18/R45*. URL: http://r-forge.rproject.org/projects/splits/.

Fayyaz, I. et al. (2022). Iqbalia Kashmirensis Gen. et Sp. Nov. from Pakistan (Teloschistaceae, Lichenized Ascomycetes) Based on Multigene Phylogeny. *Mycological Progress* 21(8):68. DOI: 10.1007/s11557-022-01823-y.

Felsenstein, J. (1973). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist* 164(6). URL: https://www.journals.uchicago.edu/doi/abs/10.1086/426002.

Fisher, M. C. et al. (2018). Worldwide Emergence of Resistance to Antifungal Drugs Challenges Human Health and Food Security. *Science* 360(6390):739–742. DOI: 10.1126/science.aap7999.

Fleming, A. (1929). On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzae. *British]ournal of Experimental Pathology,* 10:226–236.

Flynn, J. M. et al. (2020). RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proceedings of the National Academy of Sciences of the United States of America* 117(17):9451–9457. DOI: 10.1073/pnas.1921046117. pmid: 32300014.

Freire-Rallo, S. et al. (2023). To Explore Strange New Worlds – The Diversification in Tremella Caloplacae Was Linked to the Adaptive Radiation of the Teloschistaceae. *Molecular Phylogenetics and Evolution* 180:107680. DOI: 10.1016/j.ympev.2022.107680.

Frolov, I. et al. (2020). Phylogenetic Relationships within Pyrenodesmia Sensu Lato and the Role of Pigments in Its Taxonomic Interpretation. *Journal of Systematics and Evolution* 59(3):454–474. DOI: 10.1111/jse.12717.

Fujisawa, T. and Barraclough, T. G. (2013). Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Systematic Biology* 62(5):707–724. DOI: 10.1093/sysbio/syt033. pmid: 23681854.

Gardiner, D. M., Jarvis, R. S., and Howlett, B. J. (2005). The ABC Transporter Gene in the Sirodesmin Biosynthetic Gene Cluster of Leptosphaeria Maculans Is Not Essential for Sirodesmin Production but Facilitates Self-Protection. *Fungal Genetics and Biology* 42(3):257–263. DOI: 10.1016/j.fgb.2004.12.001. pmid: 15707846.

Gauslaa, Y. and Ustvedt, E. M. (2003). Is Parietin a UV-B or a Blue-Light Screening Pigment in the Lichen Xanthoria Parietina? *Photochemical and Photobiological Sciences* 2(4):424–432. DOI: 10.1039/b212532c.

Gaya, E. et al. (2003). Phylogenetic Study of Fulgensia and Allied Caloplaca and Xanthoria Species (Teloschistaceae, Lichen-Forming Ascomycota). *American Journal of Botany* 90(7):1095–1103. DOI: 10.3732/ajb.90.7.1095.

Gaya, E. et al. (2008). Phylogenetic Reassessment of the Teloschistaceae (Lichen-Forming Ascomycota, Lecanoromycetes). *Mycological Research* 112(5):528–546. DOI: 10.1016/j.mycres.2007.11.005. pmid: 18406120.

Gaya, E. et al. (2012). Implementing a Cumulative Supermatrix Approach for a Comprehensive Phylogenetic Study of the Teloschistales (Pezizomycotina, Ascomycota). *Molecular Phylogenetics and Evolution* 63(2):374–387. DOI: 10.1016/j.ympev.2012.01.012. pmid: 22306043.

Gaya, E. et al. (2015). The Adaptive Radiation of Lichen-Forming Teloschistaceae Is Associated with Sunscreening Pigments and a Bark-to-Rock Substrate Shift. *Proceedings of the National Academy of Sciences of the United States of America* 112(37):11600–11605. DOI: 10.1073/pnas.1507072112. pmid: 26324894.

Gerasimova, J. V. et al. (2022). High Diversity of Type I Polyketide Genes in Bacidia Rubella as Revealed by the Comparative Analysis of 23 Lichen Genomes. *Journal of Fungi* 8(5). DOI: 10.3390/jof8050449.

Gifford, D. R. et al. (2018). Identifying and Exploiting Genes That Potentiate the Evolution of Antibiotic Resistance. *Nature Ecology & Evolution* 2(6) (6):1033–1039. DOI: 10.1038/s41559-018-0547-x.

Gilchrist, C. L. M. et al. (2021). Cblaster: A Remote Search Tool for Rapid Identification and Visualization of Homologous Gene Clusters. *Bioinformatics Advances* 1(1):vbab016. DOI: 10.1093/bioadv/vbab016.

Gill, H., Sorensen, J. L., and Collemare, J. (2023). "Lichen Fungal Secondary Metabolites: Progress in the Genomic Era Toward Ecological Roles in the Interaction". In: *Plant Relationships: Fungal-Plant Interactions*. Ed. by B. Scott and C. Mesarich. The Mycota. Cham: Springer International Publishing:pp. 185–208. ISBN: 978-3-031-16503-0. DOI: 10.1007/978-3-031-16503-0_7.

Gluck-Thaler, E. et al. (2020). The Architecture of Metabolism Maximizes Biosynthetic Diversity in the Largest Class of Fungi. *Molecular Biology and Evolution* 37(10):2838–2856. DOI: 10.1093/molbev/msaa122.

Goering, A. W. et al. (2016). Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS Central Science*. DOI: 10.1021/acscentsci.5b00331.

Gonçalves, M. F. M. et al. (2022). Genomic and Metabolomic Analyses of the Marine Fungus Emericellopsis Cladophorae: Insights into Saltwater Adaptability Mechanisms and Its Biosynthetic Potential. *Journal of Fungi* 8(1) (1):31. DOI: 10.3390/jof8010031.

Greshake Tzovaras, B. et al. (2020). What Is in Umbilicaria Pustulata? A Metagenomic Approach to Reconstruct the Holo-Genome of a Lichen. *Genome Biology and Evolution* 12(4):309–324. DOI: 10.1093/gbe/evaa049.

Grewe, F. et al. (2020). Using Target Enrichment Sequencing to Study the Higher-Level Phylogeny of the Largest Lichen-Forming Fungi Family: Parmeliaceae (Ascomycota). *IMA Fungus* 11(1):27. DOI: 10.1186/s43008-020-00051-x.

Griffiths, S. et al. (2016). Elucidation of Cladofulvin Biosynthesis Reveals a Cytochrome P450 Monooxygenase Required for Anthraquinone Dimerization. *Proceedings of the National Academy of Sciences* 113(25):6851 LP –6856. DOI: 10.1073/pnas.1603528113.

Grigoriev, I. V. et al. (2014). MycoCosm Portal: Gearing up for 1000 Fungal Genomes. *Nucleic Acids Research* 42:699–704. DOI: 10.1093/nar/gkt1183.

Grube, M. et al. (2009). Species-Specific Structural and Functional Diversity of Bacterial Communities in Lichen Symbioses. *ISME Journal* 3(9):1105–1115. DOI: 10.1038/ismej.2009.63.

Grube, M. et al. (2013). "Lichen Genomics". In: *The Ecological Genomics of Fungi*. John Wiley & Sons, Ltd:pp. 191–212. ISBN: 978-1-118-73589-3. DOI: 10.1002/9781118735893.ch9.

Grube, M. et al. (2015). Exploring Functional Contexts of Symbiotic Sustain within Lichen-Associated Bacteria by Comparative Omics. *ISME Journal* 9(2):412–424. DOI: 10.1038/ismej.2014.138.

Gurevich, A. et al. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* 29(8):1072–1075. DOI: 10.1093/bioinformatics/btt086. pmid: 23422339.

Guzmán, G., Allen, J. W., and Gartz, J. (1998). A Worldwide Geographical Distribution of the Neurotropic Fungi, an Analysis and Discussion. *Ann. Del Mus. Civ. Di Rovereto* 14:189–280.

Haas, B. J. et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* 9(1). DOI: 10.1186/gb-2008-9-1-r7. pmid: 18190707.

Harmon, L. J. et al. (2010). Early Bursts of Body Size and Shape Evolution Are Rare in Comparative Data. *Evolution* 64(8):2385–2396. DOI: 10.1111/j.1558-5646.2010.01025.x.

Harris, B. J. et al. (2022). Divergent Evolutionary Trajectories of Bryophytes and Tracheophytes from a Complex Common Ancestor of Land Plants. *Nature Ecology & Evolution* 6(11) (11):1634–1643. DOI: 10.1038/s41559-022-01885-x.

Harvey, C. J. et al. (2018). HEx: A Heterologous Expression Platform for the Discovery of Fungal Natural Products. *Science Advances* 4(4). DOI: 10.1126/sciadv.aar5459.

Hasinoff, B. B. et al. (1997). Characterization of a Chinese Hamster Ovary Cell Line with Acquired Resistance to the Bisdioxopiperazine Dexrazoxane (ICRF-187) Catalytic Inhibitor of Topoisomerase II. *Biochemical Pharmacology* 53(12):1843–1853. DOI: 10.1016/S0006-2952(97)00013-0.

Hawksworth, D. L. and Grube, M. (2020). Lichens Redefined as Complex Ecosystems. *The New phytologist* 227(5):1281–1283. DOI: 10.1111/nph.16630.

Hebra, T. et al. (2022). Combining OSMAC, Metabolomic and Genomic Methods for the Production and Annotation of Halogenated Azaphilones and Ilicicolins in Termite Symbiotic Fungi. *Scientific Reports* 12(1) (1):17310. DOI: 10.1038/s41598-022-22256-3.

Hicks, C. et al. (2021). Evolution of the Ergot Alkaloid Biosynthetic Gene Cluster Results in Divergent Mycotoxin Profiles in Claviceps Purpurea Sclerotia. *Toxins* 13(12) (12):861. DOI: 10.3390/toxins13120861.

Higa, Y. et al. (2020). Divergence of Metabolites in Three Phylogenetically Close Monascus Species (M. Pilosus, M. Ruber, and M. Purpureus) Based on Secondary Metabolite Biosynthetic Gene Clusters. *BMC Genomics* 21(1):679. DOI: 10.1186/s12864-020-06864-9.

Higgins, C. F. (2007). Multiple Molecular Mechanisms for Multidrug Resistance Transporters. *Nature* 446(7137):749–757. DOI: 10.1038/nature05630.

Hildebrandt, U., Marsell, A., and Riederer, M. (2018). Direct Effects of Physcion, Chrysophanol, Emodin, and Pachybasin on Germination and Appressorium Formation of

the Barley (Hordeum Vulgare L.) Powdery Mildew Fungus Blumeria Graminis f. Sp. Hordei (DC.) Speer. *Journal of Agricultural and Food Chemistry* 66(13):3393–3401. DOI: 10.1021/acs.jafc.7b05977.

Hill, R., Leitch, I. J., and Gaya, E. (2021). Targeting Ascomycota Genomes: What and How Big? *Fungal Biology Reviews* 36:52–59. DOI: 10.1016/j.fbr.2021.03.003.

Hill, R. et al. (2022). Lifestyle Transitions in Fusarioid Fungi Are Frequent and Lack Clear Genomic Signatures. *Molecular Biology and Evolution* 39(4):msac085. DOI: 10.1093/molbev/msac085.

Hoang, D. T. et al. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35(2):518–522. DOI: 10.1093/molbev/msx281. pmid: 29077904.

Hof, H. (2001). Critical Annotations to the Use of Azole Antifungals for Plant Protection. *Antimicrobial Agents and Chemotherapy* 45(11):2987–2990. DOI: 10.1128/aac.45.11.2987-2990.2001.

Holm, C. et al. (1985). DNA Topoisomerase II Is Required at the Time of Mitosis in Yeast. *Cell* 41(2):553–563. DOI: 10.1016/S0092-8674(85)80028-3. pmid: 2985283.

Holman, J. D., Tabb, D. L., and Mallick, P. (2014). Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Current Protocols in Bioinformatics* 46:13.24.1–13.24.9. DOI: 10.1002/0471250953.bi1324s46. pmid: 24939128.

Honegger, R. (2023). "Lichens and Their Allies Past and Present". In: *Plant Relationships: Fungal-Plant Interactions*. Ed. by B. Scott and C. Mesarich. The Mycota. Cham: Springer International Publishing:pp. 133–183. ISBN: 978-3-031-16503-0. DOI: 10.1007/978-3-031-16503-0_6.

Humphreys, A. M. and Barraclough, T. G. (2014). The Evolutionary Reality of Higher Taxa in Mammals. *Proceedings of the Royal Society B: Biological Sciences* 281(1783). DOI: 10.1098/rspb.2013.2750. pmid: 24695424.

Humphreys, A. M. et al. (2016). Detecting Evolutionarily Significant Units above the Species Level Using the Generalised Mixed Yule Coalescent Method. *Methods in Ecology and Evolution* 7(11):1366–1375. DOI: 10.1111/2041-210X.12603.

Huneck, S. et al. (1996). *Identification of Lichen Substances*. Heidelberg/New York. DOI: 10.1007/978-3-642-85243-5_2.

Hurst, L. D., Williams, E. J., and Pál, C. (2002). Natural Selection Promotes the Conservation of Linkage of Co-Expressed Genes. *Trends in Genetics* 18(12):604–606. DOI: 10.1016/S0168-9525(02)02813-5. pmid: 12446137.

i5K Consortium (2013). The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *Journal of Heredity* 104(5):595–600. DOI: 10.1093/jhered/est050.

Jahn, L. et al. (2017). Linking Secondary Metabolites to Biosynthesis Genes in the Fungal Endophyte Cyanodermella Asteris: The Anti-Cancer Bisanthraquinone Skyrin. *Journal of Biotechnology* 257:233–239. DOI: 10.1016/j.jbiotec.2017.06.410.

Jallow, A. et al. (2021). Worldwide Aflatoxin Contamination of Agricultural Products and Foods: From Occurrence to Control. *Comprehensive Reviews in Food Science and Food Safety* 20(3):2332–2381. DOI: 10.1111/1541-4337.12734.

Jarvis, E. D. et al. (2014). Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science* 346(6215):1320–1331. DOI: 10.1126/science.1253451.

Jeffares, D. C. et al. (2015). "A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of All Protein-Coding Genes in a Genome". In: *Parasite Genomics Protocols*. Ed. by C. Peacock. Methods in Molecular Biology. New York, NY: Springer:pp. 65–90. ISBN: 978-1-4939-1438-8. DOI: 10.1007/978-1-4939-1438-8_4.

Jensen, L. H. et al. (2000). N-Terminal and Core-Domain Random Mutations in Human Topoisomerase II α Conferring Bisdioxopiperazine Resistance. *FEBS Letters* 480(2-3):201–207. DOI: 10.1016/S0014-5793(00)01934-7.

Joneson, S., Armaleo, D., and 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.

Jumper, J. et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596(7873) (7873):583–589. DOI: 10.1038/s41586-021-03819-2.

Kalyaanamoorthy, S. et al. (2017). ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nature Methods* 14:587–589. DOI: 10.1038/nmeth.4285.

Kappen, L. (2000). Some Aspects of the Great Success of Lichens in Antarctica. *Antarctic Science* 12(3):314–324. DOI: 10.1017/s0954102000000377.

Kärnefelt, I. (1989). Morphology and Phylogeny in the Teloschistales. *Cryptogamic Botany* 1:147–203.

Katoh, K., Asimenos, G., and Toh, H. (2009). "Multiple Alignment of DNA Sequences with MAFFT BT - Bioinformatics for DNA Sequence Analysis". In: ed. by D. Posada. Totowa, NJ: Humana Press:pp. 39–64. ISBN: 978-1-59745-251-9. DOI: 10.1007/978-1-59745-251-9_3.

Kealey, J. T., Craig, J. P., and Barr, P. J. (2021). Identification of a Lichen Depside Polyketide Synthase Gene by Heterologous Expression in Saccharomyces Cerevisiae. *Metabolic Engineering Communications* 13:e00172. DOI: 10.1016/j.mec.2021.e00172.

Keller, N. P. (2015). Translating Biosynthetic Gene Clusters into Fungal Armor and Weaponry. *Nature Chemical Biology* 11(9):671–677. DOI: 10.1038/nchembio.1897. pmid: 26284674.

Keuler, R. et al. (2020). Genome-Scale Data Reveal the Role of Hybridization in Lichen-Forming Fungi. *Scientific Reports*. DOI: 10.1038/s41598-020-58279-x. pmid: 32001749.

Keuler, R. et al. (2022). Interpreting Phylogenetic Conflict: Hybridization in the Most Speciose Genus of Lichen-Forming Fungi. *Molecular Phylogenetics and Evolution* 174:107543. DOI: 10.1016/j.ympev.2022.107543.

Kim, W. et al. (2021). Linking a Gene Cluster to Atranorin, a Major Cortical Substance of Lichens, through Genetic Dereplication and Heterologous Expression. *mBio* 12(3):e01111–21. DOI: 10.1128/mBio.01111-21.

Koczyk, G., Dawidziuk, A., and Popiel, D. (2015). The Distant Siblings - A Phylogenomic Roadmap Illuminates the Origins of Extant Diversity in Fungal Aromatic Polyketide Biosynthesis. *Genome Biology and Evolution* 7(11):3132–3154. DOI: 10.1093/gbe/evv204.

Kondratyuk, S. Y. et al. (2014). A Revised Taxonomy for the Subfamily Xanthorioideae (Teloschistaceae , Ascomycota). *Acta Botanica Hungarica* 56(1-2):93–123. DOI: 10.1556/ABot.56.2014.1.

Kondratyuk, S. et al. (2013). Four New Genera of Teloschistoid Lichens (Teloschistaceae, Ascomycota) Based on Molecular Phylogeny. *Acta Botanica Hungarica* 55(3-4):251–274. DOI: 10.1556/abot.55.2013.3-4.8.

Kondratyuk, S. Y. et al. (2015). Brownlielloideae, a New Subfamily in the Teloschistaceae (Lecanoromycetes, Ascomycota). *Acta Botanica Hungarica* 57(3-4):321–343. DOI: 10.1556/034.57.2015.3-4.6.

Kondratyuk, S. Y. et al. (2017). New Monophyletic Branches of the Teloschistaceae (Lichen-Forming Ascomycota) Proved by Three Gene Phylogeny. *Acta Botanica Hungarica* 59(1-2):71–136. DOI: 10.1556/034.59.2017.1-2.6.

Kondratyuk, S. Y. et al. (2018a). Coppinsiella and Seawardiella - Two New Genera of the Xanthorioideae (Teloschistaceae, Lichen-Forming Ascomycota). *Acta Botanica Hungarica* 60(3-4):369–386. DOI: 10.1556/034.60.2018.3-4.8.

Kondratyuk, S. Y. et al. (2018b). Hosseusiella and Rehmanniella, Two New Genera in the Teloschistaceae. *Acta Botanica Hungarica* 60(1-2):89–113. DOI: 10.1556/034.60.2018.1-2.7.

Kondratyuk, S. Y. et al. (2020). Three New Genera of the Teloschistaceae Proved by Three Gene Phylogeny. *Acta Botanica Hungarica* 62(1-2):109–136. DOI: 10.1556/034.62.2020.1-2.7.

Kondratyuk, S. Y. et al. (2022a). Contributions to Molecular Phylogeny of Lichens 4. New Names in the Teloschistaceae. *Acta Botanica Hungarica* 64(3-4):313–336. DOI: 10.1556/034.64.2022.3-4.7.

Kondratyuk, S. Y. and Mosyakin, S. I. (2022b). Wilketalia S.Y.Kondr., a New Name for Andina Wilk, Pabijan & Lücking, Nom. Illeg. (Teloschistaceae, Lichenized Ascomycota). *Ukrainian Botanical Journal* 79(1):3–5. DOI: 10.15407/ukrbotj79.01.003.

Kondratyuk, S. Y. et al. (2018c). Upretia, a New Caloplacoid Lichen Genus (Teloschistaceae, Lichen-Forming Ascomycota) from India. *Cryptogam Biodiversity and Assessment*(01). DOI: 10.21756/cab.esp5.

Korf, I. (2004). Gene Finding in Novel Genomes. *BMC Bioinformatics* 5. DOI: 10.1186/1471-2105-5-59. pmid: 15144565.

Kosentka, P. et al. (2013). Evolution of the Toxins Muscarine and Psilocybin in a Family of Mushroom-Forming Fungi. *PLoS ONE* 8(5):e64646. DOI: 10.1371/journal.pone.0064646. pmid: 23717644.

Kozlov, A. M. et al. (2019). RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* 35(21):4453–4455. DOI: 10.1093/bioinformatics/btz305.

Kranner, I. et al. (2008). Desiccation-Tolerance in Lichens: A Review. *The Bryologist* 111(4):576–593. DOI: 10.1639/0007-2745-111.4.576.

Kroken, S. et al. (2003). Phylogenomic Analysis of Type I Polyketide Synthase Genes in Pathogenic and Saprobic Ascomycetes. *Proceedings of the National Academy of Sciences of the United States of America* 100(26):15670–15675. DOI: 10.1073/pnas.2532165100.

Kubicek, C. P. et al. (2019). Evolution and Comparative Genomics of the Most Common Trichoderma Species. *BMC Genomics* 20(1):485. DOI: 10.1186/s12864-019-5680-7.

Kumar, S. and Hedges, S. B. (2016). Advances in Time Estimation Methods for Molecular Data. *Molecular Biology and Evolution* 33(4):863–869. DOI: 10.1093/molbev/msw026.

Laetsch, D. R. and Blaxter, M. L. (2017). BlobTools: Interrogation of Genome Assemblies. *F1000Research* 6:1287. DOI: 10.12688/f1000research.12232.1.

LaGreca, S. (1999). A Phylogenetic Evaluation of the Ramalina Americana Chemotype Complex (Lichenized Ascomycota, Ramalinaceae) Based on rDNA ITS Sequence Data. *The Bryologist* 102(4):602. DOI: 10.2307/3244250.

Lange, O. L. (1990). Twenty-Three Years of Growth Measurements on the Crustose Lichen Caloplaca a Urantia in the Central Negev Desert. *Israel Journal of Botany* 39(4-6):383–394. DOI: 10.1080/0021213X.1990.10677162.

Lawrence, D. P. et al. (2011). Interkingdom Gene Transfer of a Hybrid NPS/PKS from Bacteria to Filamentous Ascomycota. *PLoS ONE* 6(11):e28231. DOI: 10.1371/journal.pone.0028231.

Lawrey, J. D. and Diederich, P. (2003). Lichenicolous Fungi: Interactions, Evolution, and Biodiversity. *The Bryologist* 106(1):80–120. DOI: 10.1639/0007-2745(2003)106[0080:lfieab]2.0.co;2.

Le Cam, B. et al. (2019). Population Genome Sequencing of the Scab Fungal Species Venturia Inaequalis, Venturia Pirina, Venturia Aucupariae and Venturia Asperata. *G3 Genes—Genomes—Genetics* 9(8):2405–2414. DOI: 10.1534/g3.119.400047.

Le Pogam, P. et al. (2016). Spatial Mapping of Lichen Specialized Metabolites Using LDI-MSI: Chemical Ecology Issues for Ophioparma Ventosa. *Scientific Reports* 6(1)(1):37807. DOI: 10.1038/srep37807.

Leavitt, S. D. et al. (2016). Resolving Evolutionary Relationships in Lichen-Forming Fungi Using Diverse Phylogenomic Datasets and Analytical Approaches. *Scientific Reports* 6(1):222–62. DOI: 10.1038/srep22262.

Lemoine, F. et al. (2018). Renewing Felsenstein's Phylogenetic Bootstrap in the Era of Big Data. *Nature* 556(7702):452–456. DOI: 10.1038/s41586-018-0043-0.

Levy Karin, E., Mirdita, M., and Söding, J. (2020). MetaEuk—Sensitive, High-Throughput Gene Discovery, and Annotation for Large-Scale Eukaryotic Metagenomics. *Microbiome* 8(1):48. DOI: 10.1186/s40168-020-00808-x.

Li, D. et al. (2015). MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* 31(10):1674–1676. DOI: 10.1093/bioinformatics/btv033. pmid: 25609793.

Li, Y. et al. (2010a). Emodin Triggers DNA Double-Strand Breaks by Stabilizing Topoisomerase II-DNA Cleavage Complexes and by Inhibiting ATP Hydrolysis of Topoisomerase II. *Toxicological Sciences* 118(2):435–443. DOI: 10.1093/toxsci/kfq282.

Li, Y., Xu, W., and Tang, Y. (2010b). Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains. *The Journal of biological chemistry* 285(30):22764–22773. DOI: 10.1074/jbc.M110.128504.

Li, Y. et al. (2021). A Genome-Scale Phylogeny of the Kingdom Fungi. *Current Biology* 31(8):1653–1665.e5. DOI: 10.1016/j.cub.2021.01.074.

Lidgard, S. and Crane, P. R. (1988). Quantitative Analyses of the Early Angiosperm Radiation. *Nature* 331(6154) (6154):344–346. DOI: 10.1038/331344a0.

— (1990). Angiosperm Diversification and Cretaceous Floristic Trends: A Comparison of Palynofloras and Leaf Macrofloras. *Paleobiology* 16(1):77–93. DOI: 10.1017/S009483730000974X.

Lim, F. Y. et al. (2012). Genome-Based Cluster Deletion Reveals an Endocrocin Biosynthetic Pathway in Aspergillus Fumigatus. *Applied and Environmental Microbiology* 78(12):4117 LP –4125. DOI: 10.1128/AEM.07710-11.

Lind, A. L. et al. (2017). Drivers of Genetic Diversity in Secondary Metabolic Gene Clusters within a Fungal Species. *PLoS Biology* 15(11). DOI: 10.1371/journal.pbio.2003583.

Liu, A. et al. (2011). Antiproliferative and Antimetastatic Effects of Emodin on Human Pancreatic Cancer. *Oncology Reports* 26(1):81–89. DOI: 10.3892/or.2011.1257.

Liu, L. et al. (2015a). Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Scientific Reports* 5(1):10463. DOI: 10.1038/srep10463.

Liu, X.-Z. et al. (2015b). Towards an Integrated Phylogenetic Classification of the Tremellomycetes. *Studies in Mycology* 81(1):85–147. DOI: 10.1016/j.simyco.2015.12.001.

Llewellyn, T. (2019). "Revisiting Generic Concepts in the Teloschistaceae (Lichen-Forming Ascomycota) Using Molecular Phylogenetics, Algorithmic Taxa Delimitation and Molecular Networks". MSc. Queen Mary, University of London. 65 pp. URL: https://doi.org/10.34885/76.

Llewellyn, T. et al. (2023a). First Whole-Genome Sequence and Flow Cytometry Genome Size Data for the Lichen-Forming Fungus Ramalina Farinacea (Ascomycota). *Genome Biology and Evolution* 15(5):evad074. DOI: 10.1093/gbe/evad074.

Llewellyn, T. et al. (2023b). Metagenomics Shines Light on the Evolution of "Sunscreen" Pigment Metabolism in the Teloschistales (Lichen-Forming Ascomycota). *Genome Biology and Evolution* 15(2):evad002. DOI: 10.1093/gbe/evad002.

Lozano-Fernandez, J. et al. (2017). RelTime Rates Collapse to a Strict Clock When Estimating the Timeline of Animal Diversification. *Genome Biology and Evolution* 9(5):1320–1328. DOI: 10.1093/gbe/evx079.

Lücking, R. (2019). Stop the Abuse of Time! Strict Temporal Banding Is Not the Future of Rank-Based Classifications in Fungi (Including Lichens) and Other Organisms. *Critical Reviews in Plant Sciences* 38(3):199–253. DOI: 10.1080/07352689.2019.1650517.

Lücking, R., Hodkinson, B. P., and Leavitt, S. D. (2017). The 2016 Classification of Lichenized Fungi in the Ascomycota and Basidiomycota – Approaching One Thousand Genera. *The Bryologist* 119(4):361–416. DOI: 10.1639/0007-2745-119.4.361.

Lücking, R. et al. (2008). Phylogenetic Patterns of Morphological and Chemical Characters and Reproductive Mode in the Heterodermia Obscurata Group in Costa Rica (Ascomycota, Physciaceae). *Systematics and Biodiversity* 6(1):31–41. DOI: 10.1017/S1477200007002629.

Lutsak, T. et al. (2017). Testing the Correlation between Norstictic Acid Content and Species Evolution in the Cetraria Aculeata Group in Europe. *Lichenologist* 49(1):39–56. DOI: 10.1017/S0024282916000566.

Ma, X. et al. (2010). Physcion, a Natural Anthraquinone Derivative, Enhances the Gene Expression of Leaf-Specific Thionin of Barley against Blumeria Graminis. *Pest Management Science* 66(7):718–724. DOI: 10.1002/ps.1933. pmid: 20205209.

Maansson, M. et al. (2016). An Integrated Metabolomic and Genomic Mining Workflow To Uncover the Biosynthetic Potential of Bacteria. *mSystems* 1(3):e00028–15. DOI: 10.1128/mSystems.00028-15.

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two Open Source Ab Initio Eukaryotic Gene-Finders. *Bioinformatics* 20(16):2878–2879. DOI: 10.1093/bioinformatics/bth315.

Malik, M. S. et al. (2021). Journey of Anthraquinones as Anticancer Agents – a Systematic Review of Recent Literature. *RSC Advances* 11(57):35806–35827. DOI: 10.1039/D1RA05686G.

Manni, M. et al. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* 38(10):4647–4654. DOI: 10.1093/molbev/msab199.

Männle, D. et al. (2020). Comparative Genomics and Metabolomics in the Genus Nocardia. *mSystems* 5(3). DOI: 10.1128/msystems.00125-20.

McClure, R. A. et al. (2016). Elucidating the Rimosamide-Detoxin Natural Product Families and Their Biosynthesis Using Metabolite/Gene Cluster Correlations. *ACS Chemical Biology* 11(12):3452–3460. DOI: 10.1021/acschembio.6b00779.

McDonald, T. R., Gaya, E., and Lutzoni, F. (2013). Twenty-Five Cultures of Lichenizing Fungi Available for Experimental Studies on Symbiotic Systems. *Symbiosis* 59(3):165–171. DOI: 10.1007/s13199-013-0228-0.

McGary, K. L., Slot, J. C., and Rokas, A. (2013). Physical Linkage of Metabolic Genes in Fungi Is an Adaptation against the Accumulation of Toxic Intermediate Compounds. *Proceedings of the National Academy of Sciences of the United States of America* 110(28):11481–11486. DOI: 10.1073/pnas.1304461110.

McKenzie, S. K., Walston, R. F., and Allen, J. L. (2020). Complete, High-Quality Genomes from Long-Read Metagenomic Sequencing of Two Wolf Lichen Thalli Reveals Enigmatic Genome Architecture. *Genomics* 112(5):3150–3156. DOI: 10.1016/j.ygeno.2020.06.006.

Medema, M. H. et al. (2015). Minimum Information about a Biosynthetic Gene Cluster. *Nature Chemical Biology* 11(9):625–631. DOI: 10.1038/nchembio.1890.

Meiser, A. et al. (2017). Sequencing Genomes from Mixed DNA Samples - Evaluating the Metagenome Skimming Approach in Lichenized Fungi. *Scientific Reports* 7(1). DOI: 10.1038/s41598-017-14576-6.

Mello, B. et al. (2017). Fast and Accurate Estimates of Divergence Times from Big Data. *Molecular Biology and Evolution* 34(1):45–50. DOI: 10.1093/molbev/msw247.

Mesny, F. and Vannier, N. (2020). *Detecting the Effect of Biological Categories on Genomes Composition*. URL: https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition.

Miadlikowska, J. and 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(6):925–958. DOI: 10.1086/317568.

Miadlikowska, J. et al. (2014). A Multigene Phylogenetic Synthesis for the Class Lecanoromycetes (Ascomycota): 1307 Fungi Representing 1139 Infrageneric Taxa, 317 Genera and 66 Families. *Molecular Phylogenetics and Evolution* 79(1):132–168. DOI: 10.1016/j.ympev.2014.04.003.

Minh, B. Q., Hahn, M. W., and Lanfear, R. (2020). New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Molecular Biology and Evolution* 37(9):2727–2733. DOI: 10.1093/molbev/msaa106. pmid: 32365179.

Misof, B. et al. (2014). Phylogenomics Resolves the Timing and Pattern of Insect Evolution. *Science* 346(6210):763–767. DOI: 10.1126/science.1257570.

Miyauchi, S. et al. (2020). Large-Scale Genome Sequencing of Mycorrhizal Fungi Provides Insights into the Early Evolution of Symbiotic Traits. *Nature Communications* 11(1) (1):5125. DOI: 10.1038/s41467-020-18795-w.

Molnár, K. and Farkas, E. (2010). Current Results on Biological Activities of Lichen Secondary Metabolites: A Review. *Zeitschrift für Naturforschung C* 65(3-4):157–173. DOI: 10.1515/znc-2010-3-401.

Moran, N. A. and Jarvik, T. (2010). Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. *Science* 328(5978):624–627. DOI: 10.1126/science.1187113.

Mueller, S. O. et al. (1999a). Occurrence of Emodin, Chrysophanol and Physcion in Vegetables, Herbs and Liquors. Genotoxicity and Anti-genotoxicity of the Anthraquinones

and of the Whole Plants. *Food and Chemical Toxicology* 37(5):481–491. DOI: 10.1016/S0278-6915(99)00027-7.

Mueller, S. O. and Stopper, H. (1999b). Characterization of the Genotoxicity of Anthraquinones in Mammalian Cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1428(2):406–414. DOI: 10.1016/S0304-4165(99)00064-1.

Muggia, L., Schmitt, I., and Grube, M. (2008). Purifying Selection Is a Prevailing Motif in the Evolution of Ketoacyl Synthase Domains of Polyketide Synthases from Lichenized Fungi. *Mycological Research* 112(2):277–288. DOI: 10.1016/j.mycres.2007.08.020.

Muggia, L. et al. (2020). Formally Described Species Woefully Underrepresent Phylogenetic Diversity in the Common Lichen Photobiont Genus Trebouxia (Trebouxiophyceae, Chlorophyta): An Impetus for Developing an Integrated Taxonomy. *Molecular Phylogenetics and Evolution* 149:106821. DOI: 10.1016/j.ympev.2020.106821.

Murat, C. et al. (2018). Pezizomycetes Genomes Reveal the Molecular Basis of Ectomycorrhizal Truffle Lifestyle. *Nature Ecology & Evolution* 2(12) (12):1956–1965. DOI: 10.1038/s41559-018-0710-4.

Nash, T. (2008). *Lichen Biology*. Cambridge University Press. 486 pp. ISBN: 978-0-87421-656-1. DOI: 10.1007/s13398-014-0173-7.2. pmid: 15003161.

Navarro-Muñoz, J. C. et al. (2020). A Computational Framework to Explore Large-Scale Biosynthetic Diversity. *Nature Chemical Biology* 16(1):60–68. DOI: 10.1038/s41589-019-0400-9. pmid: 31768033.

Nei, M. (1967). Modification of Linkage Intensity by Natural Selection. *Genetics* 57(3):625–641. DOI: 10.1093/genetics/57.3.625. pmid: 5583732.

Nelsen, M. P. and Gargas, A. (2008). Phylogenetic Distribution and Evolution of Secondary Metabolites in the Lichenized Fungal Genus Lepraria (Lecanorales: Stereocaulaceae). *Nova Hedwigia*:115–131. DOI: 10.1127/0029-5035/2008/0086-0115.

Nelsen, M. P. et al. (2020). No Support for the Emergence of Lichens Prior to the Evolution of Vascular Plants. *Geobiology* 18(1):3–13. DOI: 10.1111/gbi.12369.

Neubauer, L. et al. (2016). Identification and Characterization of the Ergochrome Gene Cluster in the Plant Pathogenic Fungus Claviceps Purpurea. *Fungal Biology and Biotechnology* 3(1):2. DOI: 10.1186/s40694-016-0020-z.

Nguyen, L. T. et al. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32(1):268–274. DOI: 10.1093/molbev/msu300.

Nielsen, M. T. et al. (2013). Heterologous Reconstitution of the Intact Geodin Gene Cluster in Aspergillus Nidulans through a Simple and Versatile PCR Based Approach. *PLOS ONE* 8(8):e72871. DOI: 10.1371/journal.pone.0072871.

Nitiss, J. L. et al. (1994). Mutations in the gyrB Domain of Eukaryotic Topoisomerase II Can Lead to Partially Dominant Resistance to Etoposide and Amsacrine. *Molecular Pharmacology* 46(4):773–777. ISSN: 0026-895X, 1521-0111. pmid: 7969059. URL: https://molpharm.aspetjournals.org/content/46/4/773.

Nitiss, J. L. (2009). Targeting DNA Topoisomerase II in Cancer Chemotherapy. *Nature Reviews Cancer* 9(5) (5):338–350. DOI: 10.1038/nrc2607.

Noël, A. et al. (2021). Lichen-Associated Bacteria Transform Antibacterial Usnic Acid to Products of Lower Antibiotic Activity. *Phytochemistry* 181:112535. DOI: 10.1016/j.phytochem.2020.112535.

Nothias, L.-F. et al. (2020). Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nature Methods* 17(9) (9):905–908. DOI: 10.1038/s41592-020-0933-6.

Nurk, S. et al. (2017). MetaSPAdes: A New Versatile Metagenomic Assembler. *Genome Research* 27(5):824–834. DOI: 10.1101/gr.213959.116. pmid: 28298430.

Nyilasi, I. et al. (2014). Susceptibility of Clinically Important Dermatophytes against Statins and Different Statin-Antifungal Combinations. *Medical Mycology* 52(2):140–148. DOI: 10.3109/13693786.2013.828160.

Okoro, C. O. and Fatoki, T. H. (2023). A Mini Review of Novel Topoisomerase II Inhibitors as Future Anticancer Agents. *International Journal of Molecular Sciences* 24(3) (3):2532. DOI: 10.3390/ijms24032532.

Olivier-Jimenez, D. et al. (2019). A Database of High-Resolution MS/MS Spectra for Lichen Metabolites. *Scientific data* 6(1):294. DOI: 10.1038/s41597-019-0305-1.

Olivier-Jimenez, D. et al. (2021). *From Mass Spectral Features to Molecules in Molecular Networks: A Novel Workflow for Untargeted Metabolomics*. DOI: 10.1101/2021.12.21.473622. preprint.

Opanowicz, M., Blaha, J., and Grube, M. (2006). Detection of Paralogous Polyketide Synthase Genes in Parmeliaceae by Specific Primers. *Lichenologist* 38(1):47–54. DOI: 10.1017/S0024282905005529.

Orme, D. et al. (2012). *Caper: Comparative Analyses of Phylogenetics and Evolution in R.* Version 0.5. GPL-3. URL: https://cran.r-project.org/web/packages/caper/index.html.

Pagel, M. (1999). Inferring the Historical Patterns of Biological Evolution. *Nature* 401(6756) (6756):877–884. DOI: 10.1038/44766.

Palmer, J. and Stajich, J. (2019). *Nextgenusfs/Funannotate: Funannotate v1.5.3*. Zenodo. DOI: 10.5281/zenodo.2604804.

Papkou, A. et al. (2020). Efflux Pump Activity Potentiates the Evolution of Antibiotic Resistance across S. Aureus Isolates. *Nature Communications* 11(1) (1):3970. DOI: 10.1038/s41467-020-17735-y.

Paradis, E. and Schliep, K. (2019). Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics* 35(3):526–528. DOI: 10.1093/bioinformatics/bty633.

Parks, D. H. et al. (2015). CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome research* 25(7):1043–1055. DOI: 10.1101/gr.186072.114.

Parrot, D. et al. (2016). Cyaneodimycin, a Bioactive Compound Isolated from the Culture of Streptomyces Cyaneofuscatus Associated with Lichina Confinis. *European Journal of Organic Chemistry* 2016(23):3977–3982. DOI: 10.1002/ejoc.201600252.

Pavia, M. J. et al. (2023). BinaRena: A Dedicated Interactive Platform for Human-Guided Exploration and Binning of Metagenomes. *Microbiome* 11(1):186. DOI: 10.1186/s40168-023-01625-8.

Paz, G. A. de et al. (2011). Origin and Diversification of Major Clades in Parmelioid Lichens (Parmeliaceae, Ascomycota) during the Paleogene Inferred by Bayesian Analysis. *PLOS ONE* 6(12):e28161. DOI: 10.1371/journal.pone.0028161.

Pennisi, R., Ascenzi, P., and Di Masi, A. (2015). Hsp90: A New Player in DNA Repair? *Biomolecules* 5(4) (4):2589–2618. DOI: 10.3390/biom5042589.

Perlin, D. S., Rautemaa-Richardson, R., and Alastruey-Izquierdo, A. (2017). The Global Problem of Antifungal Resistance: Prevalence, Mechanisms, and Management. *The*

Lancet Infectious Diseases 17(12):e383–e392. DOI: 10.1016/S1473-3099(17)30316-X. pmid: 28774698.

Pichersky, E. and Lewinsohn, E. (2011). Convergent Evolution in Plant Specialized Metabolism. *Annual Review of Plant Biology* 62(1):549–566. DOI: 10.1146/annurev-arplant-042110-103814. pmid: 21275647.

Pizarro, D. et al. (2018). Phylogenomic Analysis of 2556 Single-Copy Protein-Coding Genes Resolves Most Evolutionary Relationships for the Major Clades in the Most Diverse Group of Lichen-Forming Fungi. *Fungal Diversity* 92(1):31–41. DOI: 10.1007/s13225-018-0407-7.

Pizarro, D. et al. (2020). Genome-Wide Analysis of Biosynthetic Gene Cluster Reveals Correlated Gene Loss with Absence of Usnic Acid in Lichen Forming Fungi. *Genome Biology and Evolution* 12(10):1858–1868. DOI: 10.1093/gbe/evaa189.

Plummer, M. et al. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6:7–11. URL: http://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf#page=7.

Pluskal, T. et al. (2010). MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC bioinformatics* 11:395. DOI: 10.1186/1471-2105-11-395. pmid: 20650010.

Prescott, T. A. K. et al. (2023). Fungal Drug Discovery for Chronic Disease: History, New Discoveries and New Approaches. *Biomolecules* 13(6) (6):986. DOI: 10.3390/biom13060986.

Prieto, M. and Wedin, M. (2017). Phylogeny, Taxonomy and Diversification Events in the Caliciaceae. *Fungal Diversity* 82(1):221–238. DOI: 10.1007/s13225-016-0372-y.

Prieto, M. and Wedin, M. (2013). Dating the Diversification of the Major Lineages of Ascomycota (Fungi). *PLOS ONE* 8(6):e65576. DOI: 10.1371/journal.pone.0065576.

Pryszcz, L. P. and Gabaldón, T. (2016). Redundans: An Assembly Pipeline for Highly Heterozygous Genomes. *Nucleic Acids Research* 44(12):e113. DOI: 10.1093/nar/gkw294.

Puttick, M. N. (2019). MCMCtreeR: Functions to Prepare MCMCtree Analyses and Visualize Posterior Ages on Trees. *Bioinformatics* 35(24):5321–5322. DOI: 10.1093/bioinformatics/btz554.

Quintanilha-Peixoto, G. et al. (2022). Phylogenomics and Gene Selection in Aspergillus Welwitschiae: Possible Implications in the Pathogenicity in Agave Sisalana. *Genomics* 114(6):110517. DOI: 10.1016/j.ygeno.2022.110517.

R Development Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D. L. (2010). Primary Controls on Species Richness in Higher Taxa. *Systematic Biology* 59(6):634–645. DOI: 10.1093/sysbio/syq060.

Ramos, D. et al. (2021). *Teloschistes Peruensis (Amended Version of 2020 Assessment). The IUCN Red List of Threatened Species 2021: E.T175710031A200461575.* URL: https://dx.doi.org/10.2305/IUCN.UK.2021-2.RLTS.T175710031A200461575.en.

Rannala, B. and Yang, Z. (2007). Inferring Speciation Times under an Episodic Molecular Clock. *Systematic Biology* 56(3):453–466. DOI: 10.1080/10635150701420643.

Resl, P. et al. (2022). Large Differences in Carbohydrate Degradation and Transport Potential among Lichen Fungal Symbionts. *Nature Communications* 13(1):2634. DOI: 10.1038/s41467-022-30218-6.

Revell, L. J. (2012). Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things). *Methods in Ecology and Evolution* 3(2):217–223. DOI: 10.1111/j.2041-210X.2011.00169.x.

Revell, L. J. and Harmon, L. J. (2022). *Phylogenetic Comparative Methods in R*. Princeton, New Jersey: Princeton University Press. 440 pp. ISBN: 978-0-691-21903-5. URL: https://press.princeton.edu/books/paperback/9780691219035/phylogenetic-comparative-methods-in-r.

Rhie, A. et al. (2021). Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species. *Nature* 592(7856) (7856):737–746. DOI: 10.1038/s41586-021-03451-0.

Robey, M. T. et al. (2021). An Interpreted Atlas of Biosynthetic Gene Clusters from 1,000 Fungal Genomes. *Proceedings of the National Academy of Sciences* 118(19):e2020230118. DOI: 10.1073/pnas.2020230118.

Rohlfs, M. (2015). Fungal Secondary Metabolite Dynamics in Fungus–Grazer Interactions: Novel Insights and Unanswered Questions. *Frontiers in Microbiology* 5:788. DOI: 10.3389/fmicb.2014.00788.

Rokas, A., Wisecaver, J. H., and Lind, A. L. (2018). The Birth, Evolution and Death of Metabolic Gene Clusters in Fungi. *Nature Reviews Microbiology* 16(12):731–744. DOI: 10.1038/s41579-018-0075-3.

Rouxel, T. et al. (2011). Effector Diversification within Compartments of the Leptosphaeria Maculans Genome Affected by Repeat-Induced Point Mutations. *Nature Communications* 2:202. DOI: 10.1038/ncomms1189. pmid: 21326234.

Ruiz-Dueñas, F. J. et al. (2021). Genomic Analysis Enlightens Agaricales Lifestyle Evolution and Increasing Peroxidase Diversity. *Molecular Biology and Evolution* 38(4):1428–1446. DOI: 10.1093/molbev/msaa301.

Saary, P., Mitchell, A. L., and Finn, R. D. (2020). Estimating the Quality of Eukaryotic Genomes Recovered from Metagenomic Analysis with EukCC. *Genome Biology* 21(1):244. DOI: 10.1186/s13059-020-02155-4.

Sabourin, M. et al. (1998). A Mutant Yeast Topoisomerase II (top2G437S) with Differential Sensitivity to Anticancer Drugs in the Presence and Absence of ATP. *Journal of Biological Chemistry* 273(44):29086–29092. DOI: 10.1074/jbc.273.44.29086. pmid: 9786915.

Salisbury, B. A. and Kim, J. (2001). Ancestral State Estimation and Taxon Sampling Density. *Systematic Biology* 50(4):557–564. DOI: 10.1080/10635150119819.

Sanders, W. B. and Masumoto, H. (2021). Lichen Algae: The Photosynthetic Partners in Lichen Symbioses. *The Lichenologist* 53(5):347–393. DOI: 10 . 1017 / s0024282921000335.

Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution* 19(1):101–109. DOI: 10.1093/oxfordjournals.molbev.a003974.

Sanmartíin, I. and Ronquist, F. (2004). Southern Hemisphere Biogeography Inferred by Event-Based Models: Plant versus Animal Patterns. *Systematic Biology* 53(2):216–243. DOI: 10.1080/10635150490423430. JSTOR: 4135409.

Santesson, J. (1970). Anthraquinones in Caloplaca. *Phytochemistry* 9(10):2149–2166. DOI: 10.1016/S0031-9422(00)85380-7.

Sayyari, E. and Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular biology and evolution* 33(7):1654–1668. DOI: 10.1093/molbev/msw079.

Scesa, P. D., Lin, Z., and Schmidt, E. W. (2022). Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nature Chemical Biology* 18(6) (6):659–663. DOI: 10.1038/s41589-022-01027-1.

Scharf, D. H. et al. (2010). Transannular Disulfide Formation in Gliotoxin Biosynthesis and Its Role in Self-Resistance of the Human Pathogen Aspergillus Fumigatus. *Journal of the American Chemical Society* 132(29):10136–10141. DOI: 10.1021/ja103262m.

Schmidt, B. H., Osheroff, N., and Berger, J. M. (2012). Structure of a Topoisomerase II–DNA–Nucleotide Complex Reveals a New Control Mechanism for ATPase Activity. *Nature Structural & Molecular Biology* 19(11) (11):1147–1154. DOI: 10.1038/nsmb.2388.

Schmidt, R. et al. (2018). Deciphering the Genome and Secondary Metabolome of the Plant Pathogen Fusarium Culmorum. *FEMS Microbiology Ecology* 94(6):fiy078. DOI: 10.1093/femsec/fiy078.

Schmitt, I. and Lumbsch, H. T. (2004). Molecular Phylogeny of the Pertusariaceae Supports Secondary Chemistry as an Important Systematic Character Set in Lichen-Forming Ascomycetes. *Molecular Phylogenetics and Evolution* 33(1):43–55. DOI: 10.1016/j.ympev.2004.04.014.

— (2009). Ancient Horizontal Gene Transfer from Bacteria Enhances Biosynthetic Capabilities of Fungi. *PLoS ONE* 4(2). DOI: 10.1371/journal.pone.0004437.

Schmitt, I. et al. (2005). Diversity of Non-Reducing Polyketide Synthase Genes in the Pertusariales (Lichenized Ascomycota): A Phylogenetic Perspective. *Phytochemistry* 66 (11 SPEC. ISS.):1241–1253. DOI: 10.1016/j.phytochem.2005.04.014. pmid: 15927215.

Sela, I. et al. (2015). GUIDANCE2: Accurate Detection of Unreliable Alignment Regions Accounting for the Uncertainty of Multiple Parameters. *Nucleic Acids Research* 43(W1):W7–W14. DOI: 10.1093/nar/gkv318.

Shen, W. et al. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* 11:e0163962. DOI: 10.1371/journal.pone.0163962. pmid: 27706213.

Shen, X.-X. et al. (2020). Genome-Scale Phylogeny and Contrasting Modes of Genome Evolution in the Fungal Phylum Ascomycota. *Science Advances* 6(45):eabd0079. DOI: 10.1126/sciadv.abd0079.

Shrestha, G. and St. Clair, L. L. (2013). Lichens: A Promising Source of Antibiotic and Anticancer Drugs. *Phytochemistry Reviews* 12(1):229–244. DOI: 10.1007/s11101-013-9283-7.

Si, H. et al. (2023). The Effects of Co-Culture on the Expression of Selected PKS Genes in the Lichenized Fungus Xanthoparmelia Taractica. *Mycological Progress* 22(6):41. DOI: 10.1007/s11557-023-01894-5.

Siddall, M. E. and Whiting, M. F. (1999). Long-Branch Abstractions. *Cladistics* 15(1):9–24. DOI: 10.1111/j.1096-0031.1999.tb00391.x.

Singh, G. et al. (2021a). Climate-Specific Biosynthetic Gene Clusters in Populations of a Lichen-Forming Fungus. *Environmental Microbiology* 23(8):4260–4275. DOI: 10.1111/1462-2920.15605.

Singh, G. et al. (2021b). Depside and Depsidone Synthesis in Lichenized Fungi Comes into Focus through a Genome-Wide Comparison of the Olivetoric Acid and Physodic Acid Chemotypes of Pseudevernia Furfuracea. *Biomolecules* 11(10). DOI: 10.3390/biom11101445.

Singh, G. et al. (2022). A Candidate Gene Cluster for the Bioactive Natural Product Gyrophoric Acid in Lichen-Forming Fungi. *Microbiology Spectrum* 0(0):e00109–22. DOI: 10.1128/spectrum.00109-22.

Slot, J. C. and Gluck-Thaler, E. (2019). Metabolic Gene Clusters, Fungal Diversity, and the Generation of Accessory Functions. *Current Opinion in Genetics and Development* 58–59:17–24. DOI: 10.1016/j.gde.2019.07.006. pmid: 31466036.

Slot, J. C. and Rokas, A. (2010). Multiple GAL Pathway Gene Clusters Evolved Independently and by Different Mechanisms in Fungi. *Proceedings of the National Academy of Sciences* 107(22):10136–10141. DOI: 10.1073/pnas.0914418107.

Smit, A. F. A., Hubley, R., and Grenn, P. (2015). *RepeatMasker Open-4.0*. RepeatMasker Open. URL: http://www.repeatmasker.org.

Søchting, U. (1997). Two Major Anthraquinone Chemosyndromes in Teloschistaceae. *Bibliotheca Lichenologica* 68:135–144.

— (2001). Chemosyndromes with Chlorinated Anthraquinones in the Lichen Genus Caloplaca. *Bibliotheca Lichenologica* 78:395–404.

Søchting, U. and Frödén, P. (2002). Chemosyndromes in the Lichen Genus Teloschistes (Teloschistaceae, Lecanorales). *Mycological Progress* 1(3):257–266. DOI: 10.1007/s11557-006-0023-x.

Søchting, U. and Lutzoni, F. (2003). Molecular Phylogenetic Study at the Generic Boundary between the Lichen-Forming Fungi Caloplaca and Xanthoria (Ascomycota, Teloschistaceae). *Mycological Research* 107(11):1266–1276. DOI: 10.1017/S0953756203008529.

Søchting, U. et al. (2021). The Lichen Genus Villophora (Teloschistaceae, Ascomycota). *The Lichenologist* 53(3):245–255. DOI: 10.1017/S0024282921000141.

Soldatou, S. et al. (2021). Comparative Metabologenomics Analysis of Polar Actinomycetes. *Marine Drugs* 19(2). DOI: 10.3390/md19020103.

Solhaug, K. A. and Gauslaa, Y. (2004). Photosynthates Stimulate the UV-B Induced Fungal Anthraquinone Synthesis in the Foliose Lichen Xanthoria Parietina. *Plant, Cell and Environment* 27(2):167–176. DOI: 10.1111/j.1365-3040.2003.01129.x.

Solhaug, K. A. and Gauslaa, Y. (1996). Parietin, a Photoprotective Secondary Product of the Lichen Xanthoria Parietina. *Oecologia* 108(3):412–418. DOI: 10.1007/BF00333715.

Solhaug, K. A., Larsson, P., and Gauslaa, Y. (2010). Light Screening in Lichen Cortices Can Be Quantified by Chlorophyll Fluorescence Techniques for Both Reflecting and Absorbing Pigments. *Planta* 231(5):1003–1011. DOI: 10.1007/s00425-010-1103-3.

Spribille, T. et al. (2016). Basidiomycete Yeasts in the Cortex of Ascomycete Macrolichens. *Science* 353(6298):488–492. DOI: 10.1126/science.aaf8287. pmid: 27445309.

Spribille, T. et al. (2022). Evolutionary Biology of Lichen Symbioses. *New Phytologist* 234:1566–1582. DOI: 10.1111/nph.18048.

Stanke, M. et al. (2006). Gene Prediction in Eukaryotes with a Generalized Hidden Markov Model That Uses Hints from External Sources. *BMC Bioinformatics* 7:62. DOI: 10.1186/1471-2105-7-62. pmid: 16469098.

Steenwyk, J. L. et al. (2019). A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera Aspergillus and Penicillium. *mBio*. DOI: 10.1128/mbio.00925-19.

Stergiopoulos, I., Zwiers, L.-H., and De Waard, M. A. (2002). Secretion of Natural and Synthetic Toxic Compounds from Filamentous Fungi by Membrane Transporters of the

ATP-binding Cassette and Major Facilitator Superfamily. *European Journal of Plant Pathology* 108(7):719–734. DOI: 10.1023/A:1020604716500.

Stocker-Wörgötter, E. (2008). Metabolic Diversity of Lichen-Forming Ascomycetous Fungi: Culturing, Polyketide and Shikimate Metabolite Production, and PKS Genes. *Natural Product Reports* 25(1):188–200. DOI: 10.1039/b606983p. pmid: 18250902.

Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* 28(6):1102–1104. DOI: 10.2144/00286ir01.

Strosnider, H. et al. (2006). Workgroup Report: Public Health Strategies for Reducing Aflatoxin Exposure in Developing Countries. *Environmental Health Perspectives* 114(12):1898–1903. DOI: 10.1289/ehp.9302.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Research* 34 (suppl_2):W609–W612. DOI: 10.1093/nar/gkl315.

Szewczyk, E. et al. (2008). Identification and Characterization of the Asperthecin Gene Cluster of Aspergillus Nidulans. *Applied and Environmental Microbiology* 74(24):7607–7612. DOI: 10.1128/AEM.01743-08.

Tagirdzhanova, G. et al. (2021). Predicted Input of Uncultured Fungal Symbionts to a Lichen Symbiosis from Metagenome-Assembled Genomes. *Genome Biology and Evolution* 13(4). DOI: 10.1093/gbe/evab047.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* 38(7):3022–3027. DOI: 10.1093/molbev/msab120.

Tamura, K. et al. (2012). Estimating Divergence Times in Large Molecular Phylogenies. *Proceedings of the National Academy of Sciences* 109(47):19333–19338. DOI: 10.1073/pnas.1213199109.

Tao, Q. et al. (2019). A Machine Learning Method for Detecting Autocorrelation of Evolutionary Rates in Large Phylogenies. *Molecular Biology and Evolution* 36(4):811–824. DOI: 10.1093/molbev/msz014.

Tavares, S. et al. (2014). Genome Size Analyses of Pucciniales Reveal the Largest Fungal Genomes. *Frontiers in Plant Science* 5. DOI: 10.3389/fpls.2014.00422.

Tehler, A. and Källersjö, M. (2001). Parmeliopsis Ambigua and P. Hyperopta (Parmeliaceae): Species or Chemotypes? *Lichenologist* 33(5):403–408. DOI: 10.1006/lich.2001.0342.

Testa, A. C., Oliver, R. P., and Hane, J. K. (2016). OculterCut: A Comprehensive Survey of at-Rich Regions in Fungal Genomes. *Genome Biology and Evolution* 8(6):2044–2064. DOI: 10.1093/gbe/evw121.

Theobald, S. et al. (2018). Uncovering Secondary Metabolite Evolution and Biosynthesis Using Gene Cluster Networks and Genetic Dereplication. *Scientific Reports* 8(1). DOI: 10.1038/s41598-018-36561-3. pmid: 30560908.

Throckmorton, K. et al. (2016). Redundant Synthesis of a Conidial Polyketide by Two Distinct Secondary Metabolite Clusters in Aspergillus Fumigatus. *Environmental Microbiology* 18(1):246–259. DOI: 10.1111/1462-2920.13007.

Tikkanen, L., Matsushima, T., and Natori, S. (1983). Mutagenicity of Anthraquinones in the Salmonella Preincubation Test. *Mutation Research/Genetic Toxicology*. DOI: 10.1016/0165-1218(83)90067-8. pmid: 6339896.

To, T.-H. et al. (2016). Fast Dating Using Least-Squares Criteria and Algorithms. *Systematic Biology* 65(1):82–97. DOI: 10.1093/sysbio/syv068.

Tolley, K. A., Chase, B. M., and Forest, F. (2008). Speciation and Radiations Track Climate Transitions since the Miocene Climatic Optimum: A Case Study of Southern African Chameleons. *Journal of Biogeography* 35(8):1402–1414. DOI: 10.1111/j.1365-2699.2008.01889.x.

Tyson, G. W. et al. (2004). Community Structure and Metabolism through Reconstruction of Microbial Genomes from the Environment. *Nature* 428(6978) (6978):37–43. DOI: 10.1038/nature02340.

U'Ren, J. M. et al. (2010). Community Analysis Reveals Close Affinities Between Endophytic and Endolichenic Fungi in Mosses and Lichens. *Microbial Ecology* 60(2):340–353. DOI: 10.1007/s00248-010-9698-2.

Urban, M., Bhargava, T., and Hamer, J. E. (1999). An ATP-driven Efflux Pump Is a Novel Pathogenicity Factor in Rice Blast Disease. *The EMBO Journal* 18(3):512–521. DOI: 10.1093/emboj/18.3.512.

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a Flexible Pipeline for Genome-Resolved Metagenomic Data Analysis. *Microbiome* 6:158. DOI: 10.1186/s40168-018-0541-1. pmid: 30219103.

Varga, T. et al. (2019). Megaphylogeny Resolves Global Patterns of Mushroom Evolution. *Nature Ecology and Evolution* 3(4):668–678. DOI: 10.1038/s41559-019-0834-1. pmid: 30886374.

Vesth, T. C. et al. (2018). Investigation of Inter- and Intraspecies Variation through Genome Sequencing of Aspergillus Section Nigri. *Nature Genetics* 50(12):1688–1695. DOI: 10.1038/s41588-018-0246-1. pmid: 30349117.

Vondrák, J. et al. (2012). Two New European Species from the Heterogeneous Caloplaca Holocarpa Group (Teloschistaceae). *The Lichenologist* 44(1):73–89. DOI: 10.1017/S0024282911000636.

Wang, M. et al. (2016). Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 34(8):828–837. DOI: 10.1038/nbt.3597.

Wang, Y. et al. (2014). Isolation and Characterization of a Polyketide Synthase Gene Cluster from Usnea Longissima. *Acta Microbiologica Sinica* 54:770–777.

Wasil, Z. et al. (2013). One Pathway, Many Compounds: Heterologous Expression of a Fungal Biosynthetic Pathway Reveals Its Intrinsic Potential for Diversity. *Chemical Science* 4(10):3845–3856. DOI: 10.1039/C3SC51785C.

Wessel, I. et al. (1999). Human Small Cell Lung Cancer NYH Cells Selected for Resistance to the Bisdioxopiperazine Topoisomerase II Catalytic Inhibitor ICRF-187 Demonstrate a Functional R162Q Mutation in the Walker A Consensus ATP Binding Domain of the α Isoform1. *Cancer Research* 59(14):3442–3450. ISSN: 0008-5472.

Wessel, I. et al. (2002). Human Small Cell Lung Cancer NYH Cells Resistant to the Bisdioxopiperazine ICRF-187 Exhibit a Functional Dominant Tyr165Ser Mutation in the Walker A ATP Binding Site of Topoisomerase IIα. *FEBS Letters* 520(1):161–166. DOI: 10.1016/S0014-5793(02)02805-3.

West, P. T. et al. (2018). Genome-Reconstruction for Eukaryotes from Complex Natural Microbial Communities. *Genome Research* 28:569–580. DOI: 10.1101/gr.228429.117. pmid: 29496730.

Westendorf, J. et al. (1990). Genotoxicity of Naturally Occurring Hydroxyanthraquinones. *Mutation Research/Genetic Toxicology*. DOI: 10.1016/0165-1218(90)90002-J. pmid: 2294411.

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Use R! 2197-5744. Switzerland: Springer Cham. 260 pp. ISBN: 978-3-319-24277-4. URL: https://doi.org/10.1007/978-3-319-24277-4.

Widhelm, T. J. et al. (2019). Multiple Historical Processes Obscure Phylogenetic Relationships in a Taxonomically Difficult Group (Lobariaceae, Ascomycota). *Scientific Reports* 9(1) (1):8968. DOI: 10.1038/s41598-019-45455-x.

Widhelm, T. J. et al. (2021). Phylogenomic Reconstruction Addressing the Peltigeralean Backbone (Lecanoromycetes, Ascomycota). *Fungal Diversity* 110(1):59–73. DOI: 10.1007/s13225-021-00476-8.

Wiemann, P. et al. (2013). Prototype of an Intertwined Secondary-Metabolite Supercluster. *Proceedings of the National Academy of Sciences* 110(42):17065–17070. DOI: 10.1073/pnas.1313258110.

Wilk, K. et al. (2021). Phylogenetic Revision of South American Teloschistaceae (Lichenized Ascomycota, Teloschistales) Reveals Three New Genera and Species. *Mycologia* 113(2):278–299. DOI: 10.1080/00275514.2020.1830672.

Wilken, P. M. et al. (2020). IMA Genome - F13 Draft Genome Sequences of Ambrosiella Cleistominuta, Cercospora Brassicicola, C. Citrullina, Physcia Stellaris, and Teratosphaeria Pseudoeucalypti. *IMA Fungus* 11. DOI: 10.1186/s43008-020-00039-7.

Williams, B. et al. (2011). Tipping the Balance: Sclerotinia Sclerotiorum Secreted Oxalic Acid Suppresses Host Defenses by Manipulating the Host Redox Environment. *PLOS Pathogens* 7(6):e1002107. DOI: 10.1371/journal.ppat.1002107.

Wirth, V. (1987). *Die Flechten Baden-Württembergs*. Ed. by E. Ulmer. Stuttgart. 528 pp.

Wisecaver, J. H. and Rokas, A. (2015). Fungal Metabolic Gene Clusters-Caravans Traveling across Genomes and Environments. *Frontiers in Microbiology* 6 (MAR). DOI: 10.3389/fmicb.2015.00161.

Wisecaver, J. H., Slot, J. C., and Rokas, A. (2014). The Evolution of Fungal Metabolic Pathways. *PLoS Genetics* 10(12). DOI: 10.1371/journal.pgen.1004816.

Wisecaver, J. H. et al. (2017). A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *The Plant Cell* 29(5):944–959. DOI: 10.1105/tpc.17.00009.

Witte, T. E. et al. (2021). Apicidin Biosynthesis Is Linked to Accessory Chromosomes in Fusarium Poae Isolates. *BMC Genomics* 22(1):591. DOI: 10.1186/s12864-021-07617-y.

Witte, T. E. et al. (2022). Untargeted Metabolomics Screening Reveals Unique Secondary Metabolite Production from Alternaria Section Alternaria. *Frontiers in Molecular Biosciences* 9:1038299. DOI: 10.3389/fmolb.2022.1038299. pmid: 36504718.

Wong, S. and Wolfe, K. H. (2005). Birth of a Metabolic Gene Cluster in Yeast by Adaptive Gene Relocation. *Nature Genetics* 37(7) (7):777–782. DOI: 10.1038/ng1584.

Wood, D. E. and Salzberg, S. L. (2014). Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments. *Genome Biology* 15(3):R46. DOI: 10.1186/gb-2014-15-3-r46.

World Health Organization (WHO) (2022). *WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action*. Geneva. ISBN: 978-92-4-006024-1. URL: https://www.who.int/publications/i/item/9789240060241.

Wu, C.-C. et al. (2011). Structural Basis of Type II Topoisomerase Inhibition by the Anticancer Drug Etoposide. *Science* 333(6041):459–462. DOI: 10.1126/science.1204117.

Wu, Y. et al. (2007). Emodin-Mediated Protection from Acute Myocardial Infarction via Inhibition of Inflammation and Apoptosis in Local Ischemic Myocardium. *Life Sciences* 81(17-18):1332–1338. DOI: 10.1016/j.lfs.2007.08.040.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24(8):1586–1591. DOI: 10.1093/molbev/msm088.

Yang, Z. and Nielsen, R. (2002). Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Molecular Biology and Evolution* 19(6):908–917. DOI: 10.1093/oxfordjournals.molbev.a004148.

Yang, Z. and Rannala, B. (2006). Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. *Molecular Biology and Evolution* 23(1):212–226. DOI: 10.1093/molbev/msj024.

Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution* 22(4):1107–1118. DOI: 10.1093/molbev/msi097.

Yoshimura, I. et al. (2002). "Isolation and Culture of Lichen Photobionts and Mycobionts". In: *Protocols in Lichenology*:pp. 3–33. DOI: 10.1007/978-3-642-56359-1_1.

Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An Early Cenozoic Perspective on Greenhouse Warming and Carbon-Cycle Dynamics. *Nature* 451(7176) (7176):279–283. DOI: 10.1038/nature06588.

Zahlbruckner, A. (1898). "Flechten (Lichenes). B. Spezieller Teil". In: *Syllabus Der Pflanzenfamilien 2*. Ed. by A. Engler. Berlin:p. 45.

— (1931). *Catalogus Lichenum Universalis 7*. Leipzig: Gebrüder Borntraeger.

— (1940). *Catalogus Lichenum Universalis 10*. Leipzig: Gebrüder Borntraeger.

Zhang, C. et al. (2018). ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees. *BMC Bioinformatics* 19. DOI: 10.1186/s12859-018-2129-y.

Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. *Molecular Biology and Evolution* 22(12):2472–2479. DOI: 10.1093/molbev/msi237.

Zhang, R.-q. et al. (2022). Genomic and Metabolomic Analyses of the Medicinal Fungus Inonotus Hispidus for Its Metabolite's Biosynthesis and Medicinal Application. *Journal of Fungi* 8(12) (12):1245. DOI: 10.3390/jof8121245.

Zhang, Y. Y. et al. (2019). Upretia Squamulosa, a New Lichen Species from the Arid Valley of Jinsha-Jiang River, China. *Phytotaxa* 402(6):288–294. DOI: 10.11646/phytotaxa.402.6.3.

Appendices

Contents

A.1 Chapter 2 Supplementary Materials 193
A.2 Chapter 3 Supplementary Materials 199
A.3 Chapter 4 Supplementary Materials 219
A.4 Chapter 5 Supplementary Materials 225
A.5 *Ramalina farinacea* genome note (doi:10.1093/gbe/evad074) 227

A.1 Chapter 2 Supplementary Materials

Methods

Table A.1: Collection information for 24 newly sequenced lichens in Chapter 2.

Taxon	Locality	Legit	Collection date
Diploschistes diacapsis	Catalunya, Spain	XL, AA	04.09.2018
Umbilicaria vellea	Catalunya, Spain	XL, AA	22.09.2018
Letrouitia transgressa	Sri Lanka	AA	31.05.2017
Letrouitia transgressa (2)	Sri Lanka	AA	26.04.2017
Teloschistes flavicans	Devon, UK	TBL	16.10.2019
T. chrysophthalmus	Catalunya, Spain	XL, AA	02.04.2019
Teloschistes peruensis	Santiago, Peru	OW-1716	20.01.2015
Xanthomendoza fulva	Catalunya, Spain	XL, AA	20.09.2018
Flavoplaca oasis	Negev Desert, Israel	MT	12.11.2018
Xanthoria sp.	Telve, Italy	TBL	08.09.2018
Xanthoria sp. 2	Belluno, Italy	TBL	01.09.2019
Xanthoria aureola	Catalunya, Spain	XL, AA	No date
Xanthoria steineri	Negev Desert, Israel	MT	04.12.2018
Xanthoria mediterranea	Negev Desert, Israel	MT	05.12.2018
Gyalolechia ehrenbergii	Negev Desert, Israel	MT	12.11.2018
Letrouitia leprolyta	Udawalawa, Sri Lanka	AA	No date
Usnochroma carphinea	Catalunya, Spain	XL, AA	14.04.2019
Caloplaca ligustica	Catalunya, Spain	XL	16.04.2019
Caloplaca aetnensis	Catalunya, Spain	XL, AA	16.04.2019
Caloplaca aegea	Catalunya, Spain	XL, AA	29.10.2018
Variospora aurantia	Negev Desert, Israel	MT	04.12.2018
Seirophora villosa	Negev Desert, Israel	MT	04.12.2018
Seirophora lacunosa	Negev Desert, Israel	MT	12.11.2018
Seirophora lacunosa (2)	Catalunya, Spain	AA, XL	20.05.2019

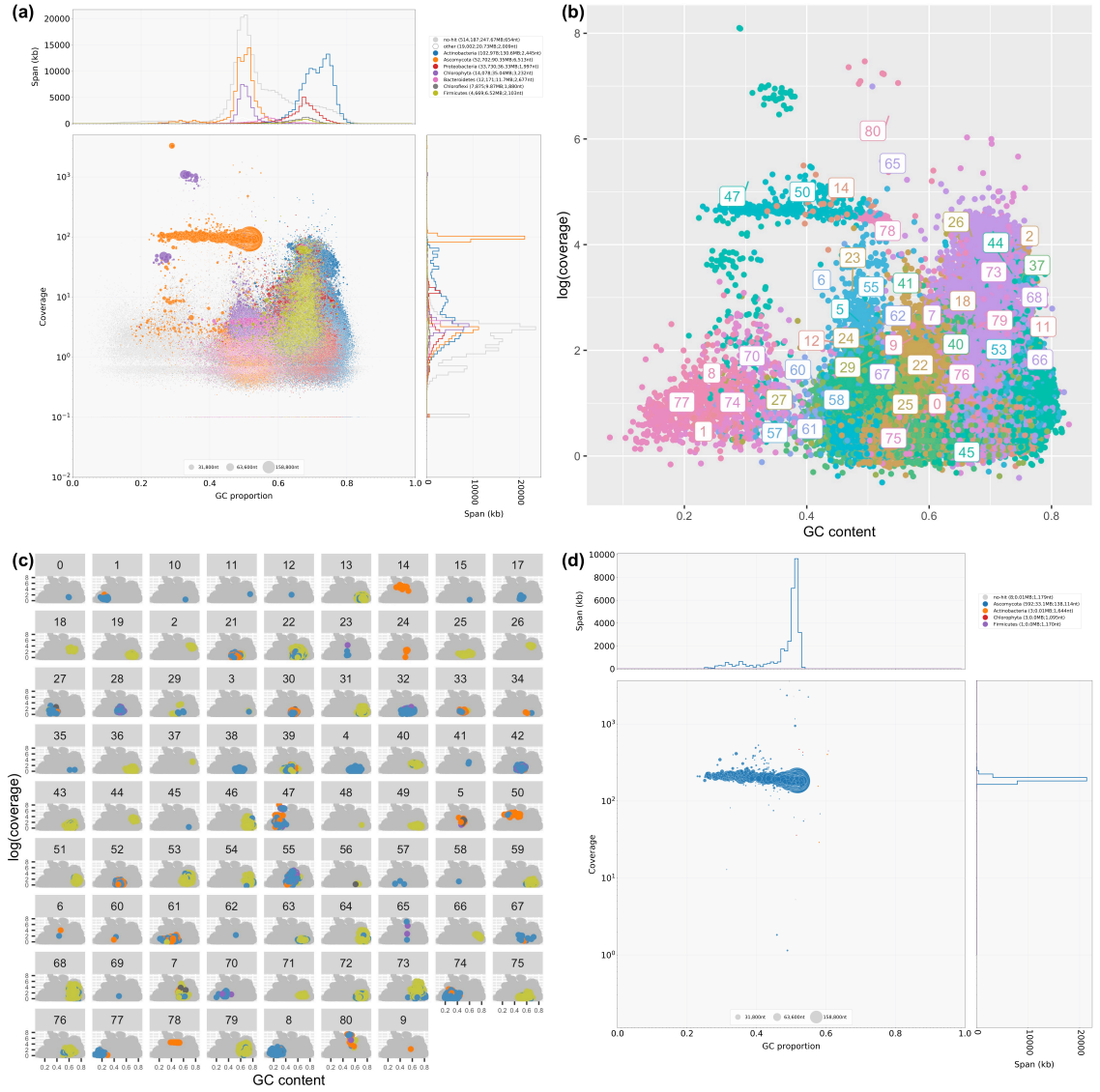


Figure A.1: An example of mycobiont reads filtering combining BlobTools and CONCOCT methods for the lichen metagenome data of *Xanthoria mediterranea* (*Teloschistaceae*, *Ascomycota*). (**a**) Blobplot pre-mycobiont filtering showing the GC proportion against coverage for each contig in the metagenome assembly. Histograms for both axes are shown. Circle colour represents taxonomic identification via BLAST/DIAMOND. Circle diameter represents contig length. Legend shows contig count, total span (Megabases) and N50 for each taxon. Light grey circles are contigs that could not be taxonomically identified. (**b**) Plot showing GC proportion against coverage for the same metagenome as in (**a**). Contigs (points) are coloured by the CONCOCT metagenome bin to which they were assigned and labelled accordingly with a unique bin number. (**c**) Facet plot showing GC proportion against coverage for each CONCOCT bin individually. Bin numbers are shown above each facet and contigs belonging to that bin are highlighted (coloured points) against the full metagenome assembly (light grey points). Points are coloured by summarised BlobTools taxonomic ID: orange = *Ascomycota*, green = Bacteria, purple = *Chlorophyta*, blue = no hit, dark grey = other. (**d**) Blobplot after merging CONCOCT bins 14, 50, 65, 78, and 80 and removing remaining non-mycobiont reads. Colour coding for each subplot is unique and not comparable between subplots.

Table A.2: Lichen genome assemblies in the custom *Lecanoromycetes* database used in BLAST, phylogenomics and comparative genomics analysis. N50 = contig N50. * = assembly from metagenomic data. Numbers in parentheses after species names indicate multiple assemblies of same taxon. † = *Teloschistaceae*

Species	**Genome size (Mbp)**	**N50 (kb)**	**Database**	**Accession number**
*Alectoria sarmentosa**	39.96	93	GenBank	GCA_009733775.1
Cladonia macilenta	37.12	1469	GenBank	GCA_000444155.1
Cladonia metacorallifera	36.68	1591	GenBank	GCA_000482085.2
Cladonia rangiferina	35.67	273	GenBank	GCA_006146055.1
Cladonia uncialis	32.85	35	GenBank	GCA_002927785.1
Cladonia grayii	34.62	243	JGI Mycocosm	Cgr/DA2myc/ss v1.0
Cyanodermella asteris	28.67	1790	GenBank	GCA_900618795.1
Evernia prunsatri	40.35	264	GenBank	GCA_003184365.1
Gyalolechia flavorubscens†	34.47	1693	GenBank	GCA_000442125.1
Lasallia hispanica	41.21	145	GenBank	GCA_003254425.1
Lasallia pustulata	32.91	1808	GenBank	GCA_008636195.1
Lasallia pustulata (2)	39.23	21	GenBank	GCA_900169345.1
Lasallia pustulata (3)	39.23	21	GenBank	GCA_000938525.1
*Letharia columbiana**	52.27	667	GenBank	GCA_014066305.1
*Letharia lupina**	49.20	2098	GenBank	GCA_014066315.1
Pseudevernia furfuracea	37.80	1178	GenBank	GCA_003184345.1
Ramalina intermedia	26.25	273	GenBank	GCA_003073195.1
Ramalina peruensis	26.99	40	GenBank	GCA_001956345.1
Rusavskia elegans†	44.32	386	GenBank	GCA_011316305.1
Usnea florida	44.32	502	JGI Mycocosm	ATCC18376 v1.0
Usnea hakonensis	41.20	166	GenBank	GCA_013423325.1
Umbilicaria muehlenbergii	34.81	7009	GenBank	GCA_000611775.1
Xanthoria parietina†	31.90	1731	JGI Mycocosm	46-1-SA22 v1.1

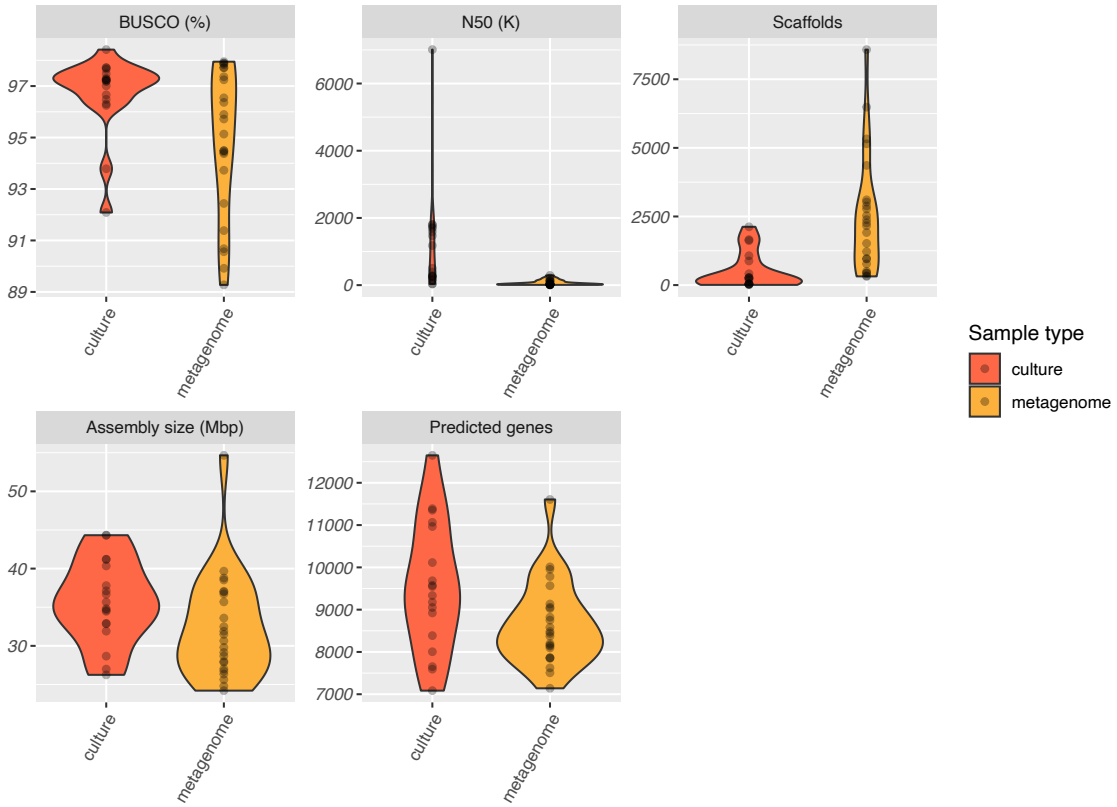


Figure A.2: Violin plots comparing post-assembly summary statistics of published lichenised fungal genomes from pure mycobiont culture (orange) with lichenised fungal genomes from metagenomic data (yellow). Top row from left to right: Percentage complete BUSCO genes, N50 score in 1000s of bases (K), Number of assembly scaffolds. Bottom row: Assembly size in megabase pairs (Mbp), Number of predicted genes.

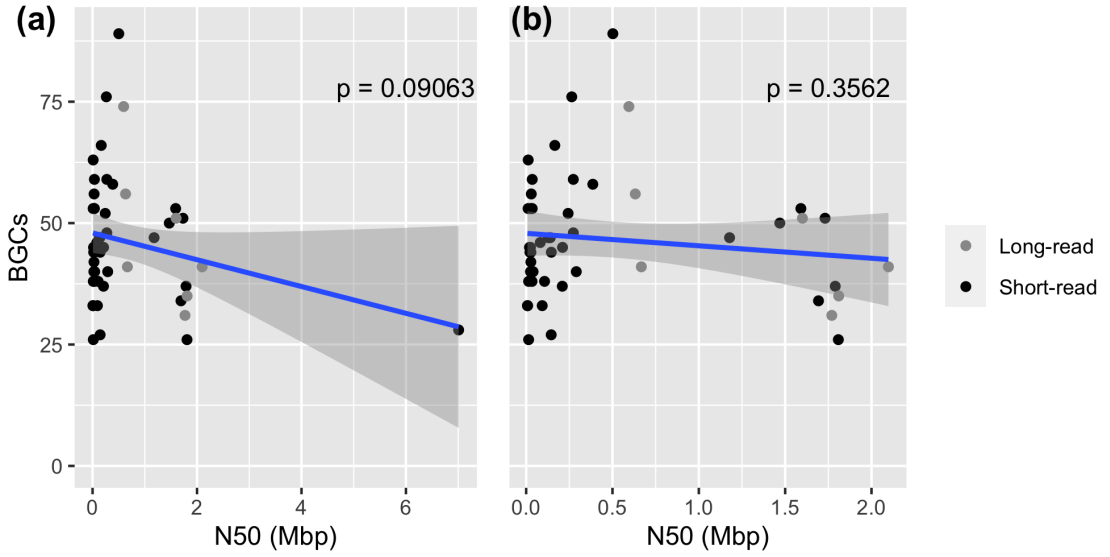


Figure A.3: Scatterplots showing the relationship between number of detected BGCs per genome and assembly fragmentation (N50) for 50 lichen mycobiont genome assemblies. Dataset includes the 45 assemblies shown in Table 2.1 and Supplementary Table A.2 plus five additional long-read genomes from the following publications (Allen et al. 2021; Wilken et al. 2020; Singh et al. 2021b; Gerasimova et al. 2022). (**a**) Number of BGCs shows no significant relationship with N50, partial F-test at a $p < 0.05$ significance level ($F_{48} = 2.9819, p = 0.09063$). (**b**) Same as (**a**) except for removal of a potential outlier *Umbilicaria muehlenbergii* which had a significantly higher N50 than other assemblies. Again, number of BGCs shows no significant relationship with N50 ($F_{47} = 0.8684, p = 0.3562$). Sequencing technology (long-read or short-read) is highlighted by dot colour. Fitted regression lines and 95% confidence intervals (coloured boxes around regression lines) were predicted using the linear model.

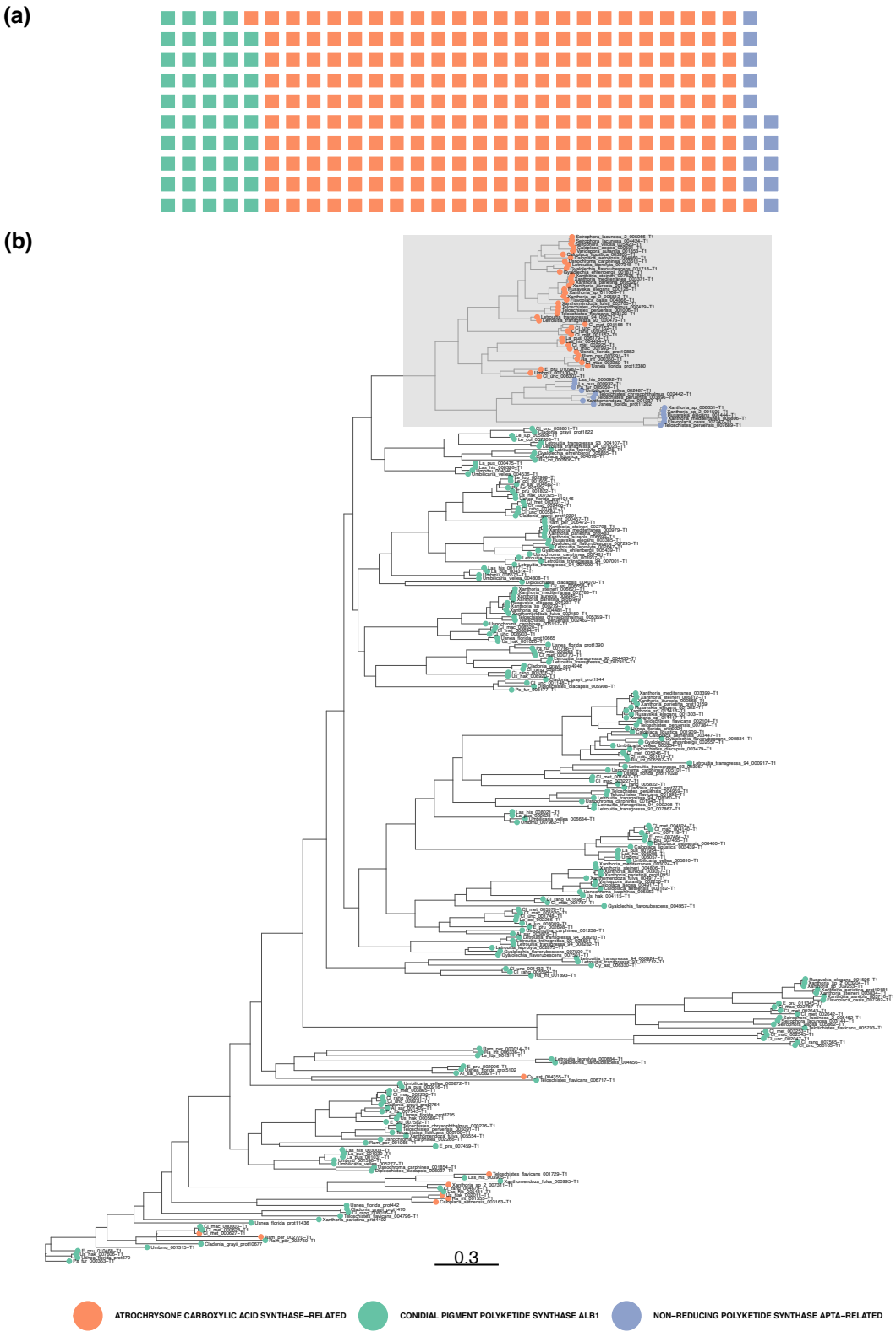


Figure A.4: Results of Interpro functional analysis performed with PANTHER to annotate the gene sequences on the orthogroup containing my putative anthraquinone PKSs. (**a**) Waffle chart representation of PKS protein PANTHER functional annotations. Each square represents a PKS protein in the orthogroup and is coloured by the assigned PANTHER function from Interpro analysis. (**b**) Maximum likelihood gene tree for the orthogroup (shown in Figure 2.3) with PANTHER functional annotation plotted as coloured tip dots. Colours follow the key in (**a**). My putative anthraquinone PKSs clade is highlighted in the grey block.

A.2 Chapter 3 Supplementary Materials

Methods

To test the impact of alignment trimming I performed maximum likelihood phylogenetic inference from both trimmed and untrimmed alignments and compared the two. For the 84 taxon genome-scale dataset I extracted single-copy orthgroups present in 50% of the taxa. I aligned orthologous sequences using MUSCLE5 (Edgar 2022) and trimmed them with trimAl v1.4.rev12 with the '—automated1' option (Capella-Gutiérrez et al. 2009). For both trimmed and untrimmed alignment, I replaced terminal gaps with '?' and concatenated all 599 loci using AMAS (Borowiec 2016). I used modeltest-ng to choose the best substitution model for each gene (Darriba et al. 2020) and produced two concatenated maximum likelihood phylogenies (trimmed and untrimmed) using RAxML-ng (**kozlov˙raxml-ng˙2019**). Bootstrapping convergence was tested using the autoMRE criterion within RAxML-NG. I compared the trimmed and untrimmed phylogenies using the tanglegram function of the dendextend R package (Supplementary Figure A.5).

Table A.3: Collection information for 59 newly sequenced lichens in Chapter 3.

Taxon	Collection information (Locality, Coll., Date)
Blastenia crenularia	Rogaland, Norway, EM, AE, 04.06.2021
Brigantiaea leucoxantha	Tamaulipas, Mexico, AH, 01.11.2020
Calogaya cf. arnoldii	Crete, Greece, JV, 08.05.2005
Calogaya cf. arnoldii	Upper Galilee, Israel, MT, 30.01.2017
Caloplaca s.l. sp. 1	Coahuila, Mexico, AH, 19.07.2020
Caloplaca s.l. sp. 2	Tamaulipas, Mexico, AH, 28.10.2020
Caloplaca s.l. sp. 3	Nuevo León, Mexico, AH, 21.02.2018
Dirinaria leoncita	Veracruz, Mexico, GG, 01.01.2017
Flavoplaca austrocitrina	Mount Carmel National Park, Israel, MT, 05.05.2021
Flavoplaca cf. flavocitrina	Atlit Beach, Israel, MT, 12.05.2021
Flavoplaca citrina	Mount Carmel National Park, Israel, MT, 05.05.2021
Flavoplaca holocarpa	Mount Carmel National Park, Israel, MT, 05.05.2021
Flavoplaca limonia	Las Nieves, Lanzarote, MLG, ED, 02.05.2021
Flavoplaca navasiana	Atlit Beach, Israel, MT, 12.05.2021
Flavoplaca navasiana	Guatisea, Lanzarote, MLG, ED, 23.01.2021
Gyalolechia fulgens	Israel, Negev desert, MT, 04.12.2018
Gyalolechia gomerana	Los Ancones, Lanzarote, MLG, ED, 01.11.2020
Gyalolechia sp. 1	Guatiza, Lanzarote, MLG, ED, 21.02.2021
Gyalolechia sp. 2	Cortijo de D. Juan Feo, Lanzarote, MLG, ED, 04.04.2021
Haloplaca sp. 1	Mount Carmel National Park, Israel, MT, 05.05.2021
Haloplaca sp. 2	Pechiguera, Lanzarote, MLG, ED, 25.04.2021
Heterodermia aff. obscurata	Veracruz, Mexico, GG, 26.08.2016
Ikaeria aurantiellina	Barranco de Elvira Sánchez, Lanzarote, MLG, ED, 21.02.2021
Kuettlingeria erythrocarpa	Mount Carmel National Park, Israel, MT, 23.05.2021
Letrouitia sp. 1	Data missing, AA, 28.04.2017
Letrouitia subvulpina	Udawalawa, Sri Lanka, AA, 04.04.2017
Loekoesia sp. 1	Nuevo León, Mexico, AH, 22.11.2020
Polycauliona sp. 1	Nuevo León, Mexico, AH, 30.08.2020
Polycauliona sp. 2	Las Nieves, Lanzarote, MLG, ED, 10.01.2021
Pyrenodesmia sp. 2	Santa Catarina, Mexico, AH, 13.06.2020
Pyrenodesmia sp. 3	Guatisea, Lanzarote, MLG, ED, 23.01.2021
Rusavskia elegans	Nuevo León, Mexico, AH, 30.08.2020
Seirophora lacunosa	Catalunya, Spain, AA, XL, 20.05.2019
Seirophora scorigena	Guatisea, Lanzarote, MLG, ED, 28.01.2021
Seirophora villosa	Las Nieves, Lanzarote, MLG, ED, 02.05.2021
Squamulea sp. 1	Nuevo León, Mexico, AH, 09.02.2019
Squamulea squamosa	Nuevo León, Mexico, AH, 28.12.2020
Tayloriellina microphyllina	Coahuila, Mexico, AH, 26.12.2020
Teloschistes cf. exilis	Nuevo León, Mexico, AH, 05.01.2020
Teloschistes chrysophthalmus	Data missing, Data missing, Data missing
Teloschistes exilis	Nuevo León, Mexico, AH, 05.01.2020
Teloschistes exilis	Data missing, Data missing, Data missing
Teloschistes flavicans	Veracruz, Mexico, GG, 01.02.2021
Teloschistes hypoglaucus	Veracruz, Mexico, GG, 01.02.2021
Variospora cf. flavescens	Montaña blanca, Lanzarote, MLG, ED, 31.01.2021
Variospora fuerteventurae	Morro Alto, Lanzarote, MLG, ED, 11.07.2021
Variospora sp. 1	Tiagua, Lanzarote, MLG, ED, 02.05.2021
Variospora velana	Switzerland, EG, PN, NH, AG, UA, US, 13.08.2006
Wetmoreana brouardii	Jalisco, Mexico, EG, 05.06.2020
Wetmoreana decipioides	Nuevo León, Mexico, AH, 22.11.2020
Wetmoreana sp. 5	Nuevo León, Mexico, AH, 22.11.2020
Xanthocarpia lactea	Mount Carmel National Park, Israel, MT, 05.05.2021
Xanthocarpia marmorata	Mount Carmel National Park, Israel, MT, 23.05.2021
Xanthocarpia ochracea	Mount Carmel National Park, Israel, MT, 12.05.2021
Xanthomendoza concinna	Coahuila, Mexico, AH, 06.07.2020
Xanthomendoza sp. 1	Coahuila, Mexico, AH, 06.08.2020
Xanthomendoza sp. 2	Nuevo León, Mexico, AH, 23.08.2020
Xanthomendoza weberi	Nuevo León, Mexico, AH, 23.08.2020
Xanthomendoza weberi	Nuevo León, Mexico, AH, 28.12.2020

Table A.4: Multilocus *Teloschistales* sequences included in 552-taxon molecular phylogenetic analysis. GenBank accession or ID numbers shown. Locality format is country, collector, voucher information. Sequences marked with 'gea' are those from Gaya et al. (unpublished), which are not yet published on GenBank. Sequences marked 'new' are those newly sequenced in Llewellyn (2019).

Species	Locality/Voucher	ITS	LSU	mrSSU	RPB1	RPB2	RPB2-2nd	MCM7	RET1
'Caloplaca' cupulifera	Australia, Vondrák 4910, CBFS	MT967374	–	–	–	–	–	–	–
'Caloplaca' floridana	Galapagos, Bungartz 7935, CDS 38445	MT967379	–	–	–	–	–	–	–
'Caloplaca' sideritis	USA, South Dakota, Adavita 5073, LD	MT967382	–	–	–	–	–	–	–
"Caloplaca" albolutescens	*Sweden, Arup L09030, LD*	KC179423	KC179502	—	—	—	—	—	—
"Caloplaca" sp. 41	*USA, Oregon, McCune 27980, ORE*	KC179430	KC179509	—	—	—	—	—	—
"Caloplaca" transcaspica	*Ukraine, Vondrák JV5466, CBFS*	EU192156	—	—	—	—	—	—	—
Amundsenia approximata	*Sweden, Arup L02084, LD*	KJ789964	KJ789974	KJ789972	—	—	—	—	—
Amundsenia austrocontinentalis	*Antartica, McMurdo Dry Valleys, Seppelt 27537, HO*	KJ789961	KJ789975	—	—	—	—	—	—
Andina citrinoides sp. nov.	*Bolivia, Wilk 7553, KRAM holotype, LPB isotype*	MT754240	–	MT754208	–	–	–	–	–
Aridoplaca peltata sp. nov.	Bolivia, Flakus 14785, KRAM	MT754232	MT741635	MT754201	–	–	–	–	–
Athallia cerinelloides	*Sweden, Arup L06208, LD (ITS); Sweden, Arup L07202, LD (LSU, SSU)*	KC179339	KC179477	KC179147	—	—	—	—	—
Athallia holocarpa	*Czech Republic, Hruby Jesenik Mts, Vondrak*	KT934392			—	—	—	—	—
Athallia holocarpa	*Sweden, Arup L04019, LD*		KC179478	KC179148	—	—	—	—	—
Austroplaca cirrochrooides	*Chile, Søchting 11300, C*	KC179082	KC179482	KC179152	—	—	—	—	—
Austroplaca darbishirei	*Antarctica, Antarctic Peninsula, Søchting 11401, C*	KC179083	KC179483	KC179153	—	—	—	—	—
Austroplaca hookeri	*Antarctica, South Shetland Isl., Søchting 7611, C*	KC179085	KC179484	KC179154	—	—	—	—	—
Austroplaca lucens	*France, Kerguelen Isl., Søchting 9417, C*	KC179087	KC179485	KC179155	—	—	—	—	—
Austroplaca soropelta	*Iceland, Frödén 650, LD (ITS); Iceland, Søchting 7536, C (LSU, SSU)*	KC179089	KC179487	KC179157	—	—	—	—	—
Austroplaca sp. 10	*Chile, Elvebakk 98:349, TROM*	KC179093	KC179490	KC179160	—	—	—	—	—
Austroplaca sp. 7	*Chile, Søchting 10293, C*	KC179090	KC179488	KC179158	—	—	—	—	—
Austroplaca sp. 9	*Chile, Søchting 11272, C*	KC179092	KC179489	KC179159	—	—	—	—	—
Blastenia circumpolaris	*Australia, VIC, Frödén 1479, LD*	KC179414	—	—	—	—	—	—	—
Blastenia ferruginea	*Spain, Søchting 9996, C*	KC179416	KC179493	KC179163	—	—	—	—	—
Brigantiaea cf leucoxantha 71	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 40691, D. & M. Brand, Sérusiaux, P.v.d. Boom Herb.*	JQ301654	JQ301477	JQ301542	gea	gea	—	gea	—
Brigantiaea cf tricolor 70	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 40492, D. & M. Brand, Sérusiaux, P.v.d. Boom Herb.*	JQ301655	—	JQ301543	gea	gea	gea	gea	gea
Brigantiaea fuscolutea 65	*USA, Alaska, Spribille 26961, GZU*	—	JQ301478	JQ301544	gea	—	gea	gea	gea
Brigantiaea leucoxantha 17	*USA, Alabama, Gaya 17, DUKE*	—	JQ301479		—	gea	—	—	—
Brigantiaea sorediata 4962	*Japan, Honshu, Kashiwadani 44106, H*	—	*JQ301480	*JQ301546	—	—	—	—	—
Brigantiaea tricolor 26	*Indonesia, West Java, Sudirman & Sipman 51473, B*	JQ301656	—	—	—	—	—	—	gea
Brownliella aequata	*Australia: Western Australia, Kondratyuk 20430, LD*	KF264627	KF264688	KF264662	—	—	—	—	—
Brownliella kobeana	*South Korea, Jayalal, Park & Ryu 120032, KoLRI 014623*	KT456212	—	KT456227	—	—	—	—	—
Brownliella montisfracti	*Australia: NSW, Cooma, Kondratyuk 20494, KW-L*	KF264625	KF264686	KF264660	—	—	—	—	—
Bryoplaca jungermanniae	*Denmark, Greenland, Søchting 10451, C*	KC179420	—	—	—	—	—	—	—
Bryoplaca sinapisperma	*Norway, Arup L08184, LD*	KC179421	KC179495	—	—	—	—	—	—
Bryoplaca tetraspora	*Antarctica, South Shetland Isl. Søchting 7979, C*	KC179422		—	—	—	—	—	—
Bryoplaca tetraspora	*Denmark, Greenland, Søchting 10480, C*		KC179496		—	—	—	—	—
C. saviczii	Kamchatka, Russia, holotype	MN814226.1	MW227509	MW227327	–	–	–	–	–
Calogaya altynis	*China, 20072284, XJU-NALH*	KY749122	—	—	—	—	—	—	—
Calogaya biatorina subsp. asiatica	*Turkey, CL1119, ERC*	KY749014	—	—	—	—	—	—	—
Calogaya haloxylonis	*China, 20080186, XJU-NALH*	KY749022	—	—	—	—	—	—	—
Calogaya mogoltanica	*Iran, Kondratyuk SK201135, KW-L*	KJ133453	KJ133510	KJ133492	—	—	—	—	—
Calogaya orientalis	*China, 20136740, XJU-NALH*	KY749063	—	—	—	—	—	—	—
Calogaya qinghaiensis	*China, Qinghai province, Oh, Han, & Hur CH140258*	KU199679	—	—	—	—	—	—	—
Calogaya xanthoriella	*China, 20081479, XJU-NALH*	KY749097	—	—	—	—	—	—	—
Calogaya xinjiangis	*China, 20136648, XJU-NALH*	KY749025	—	—	—	—	—	—	—
Caloplaca aegaea 248	*Spain, Catalunya, Gaya 248 & Llimona, BCN*	EU639597	—	—	—	gea	gea	—	—
Caloplaca aetnensis 236	*Spain, Catalunya, Llimona s.n., BCN*	EU639590	gea	—	—	—	gea	—	—
Caloplaca aff arenaria 101	*USA, California, Gaya 17 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea

Caloplaca aff arnoldii 241	*Peru, Gaya 13 & Ramos, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca aff flavorubescens 160	*Mexico, Baja California Sur, Gaya 12 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	gea
Caloplaca aff flavorubescens 164	*Mexico, Baja California Sur, Gaya 5 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca aff maritima 110	*USA, California, Gaya 8B & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	—
Caloplaca aff peliophylla 152	*Mexico, Baja California Sur, Gaya 2C & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	gea
Caloplaca alcarum 459	*Greenland, Hansen 228, C 11185*	EU639644	—	—	—	—	—	—	—
Caloplaca arnoldii ssp oblit 5	*Sweden, Gaya, Navarro-Rosinés, Arup, Søchting, Hladun & Gómez-Bolea, BCN*	gea	gea	gea	gea	gea	gea	—	gea
Caloplaca aurantia 13326	*Spain, Catalunya, Navarro-Rosinés & Llimona, BCN 13326*	EU639602	gea	—	—	gea	gea	—	—
Caloplaca bolacina 414	*Mexico, Baja California sur, Navarro-Rosinés, BCN-Lich 12532*	EU639626	gea	gea	—	gea	gea	—	gea
Caloplaca brouardii 159	*Mexico, Baja California sur, Gaya 3 & Lutzoni, DUKE*	gea	—	gea	gea	gea	gea	gea	gea
Caloplaca cancaraxiticola 375	*Spain, Castilla-La Mancha, Llimona & Egea s.n., MUB*	EU639608	gea	—	—	gea	gea	—	—
Caloplaca carphinea 51	*Spain, Catalunya, Gaya, Fernández-Brime & Llimona, BCN 13714*	EU639595	JQ301482	JQ301548	gea	gea	gea	gea	gea
Caloplaca cerina 16	*USA, Alaska, Gaya, Geml, Gomà, DUKE*	JQ301658	JQ301483	JQ301549	—	gea	—	—	—
Caloplaca cf amniospila 100	*USA, California, Gaya 13 & Lutzoni, DUKE*	gea	gea	gea	gea	—	gea	gea	—
Caloplaca cf boulyi 114	*USA, California, Gaya 4 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	—
Caloplaca cf chrysophthalma 157	*Mexico, Gaya 4 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca cf crenulatella 108	*USA,California, Gaya 21 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca cf luteoalba 155	*USA, California, Gaya 4 & Lutzoni, DUKE*	gea	gea	gea	—	—	gea	gea	gea
Caloplaca cf pyracea 99	*USA, California, Gaya 18D & Lutzoni, DUKE*	gea	gea	gea	—	gea	gea	—	gea
Caloplaca chalybaea 38	*Sweden, Gaya, Navarro-Rosinés, Arup, Søchting, Hladun & Gómez-Bolea, BCN*	JQ301659	JQ301484	JQ301550	gea	gea	gea	—	gea
Caloplaca chilensis 68	*Chile, Stanton 0212-0508, SGO*	JQ301660	JQ301485	JQ301551	gea	—	gea	gea	—
Caloplaca chlorina 191	*Sweden, Arup L09009, LD*	gea	gea	—	gea	gea	—	—	—
Caloplaca chlorina 2055	*Czech Republic, Vondrak 4307, CBFS*	—	JQ301486	—	—	—	—	—	—
Caloplaca chlorina 57	*Bulgaria, Vondrak 4307, CBFS*	164520653	—	—	—	—	gea	—	—
Caloplaca cinnabarina 149	*Mexico, Baja California, Gaya 11 & Lutzoni, DUKE*	—	—	gea	gea	gea	gea	gea	gea
Caloplaca cinnamomea 24	*Iceland, Lutzoni & Gueidan, DUKE*	—	JQ301487	JQ301552	gea	—	gea	—	gea
Caloplaca cirrochroa 13711	*Spain, Catalunya, Navarro-Rosinés, Gaya 209 & Hladun, BCN 13711*	EU639611	—	—	—	—	—	—	—
Caloplaca coralloides 79	*USA, California, Nash III 41158, ASU 511740*	gea	gea	gea	—	gea	—	—	gea
Caloplaca coronata 394	*Spain Catalunya, Gaya 394, BCN*	183585252	—	—	gea	—	gea	—	—
Caloplaca coronata 48	*Spain, Almeria, P. & B. v.d. Boom 38374, P.v.d. Boom Herb.*	—	JQ301488	JQ301553	—	gea	—	—	gea
Caloplaca crenularia 19	*Sweden, Gaya, Navarro-Rosinés, Arup, Søchting, Hladun & Gómez-Bolea, BCN*	JQ301661	JQ301489	—	gea	—	gea	—	gea
Caloplaca crenulatella 111	*USA, California, Gaya 23 & Lutzoni, DUKE*	gea	gea	—	—	—	gea	—	—
Caloplaca decipiens 2 342b	*Denmark, Søchting, C*	—	—	39752525	—	—	—	—	—
Caloplaca decipiens 2377 342b	*USA, Nebraska, Wetmore 77865, MIN*	183585258	—	—	—	—	—	—	—
Caloplaca decipioides 188	*South Korea, Gangwon Province, Thor 20768, holotype UPS, isotypes NIBR, UPS*	gea	gea	gea	gea	gea	—	—	—
Caloplaca elegantissima 75	*Namibia, Crittenden*	gea	gea	gea	gea	gea	gea	—	gea
Caloplaca eos 190	*Australia, NSW, Kärnefelt 20044701, Filson & Kondratyuk 20475, holotype CANB, isotypes PERTH, AQ, MEL, LD, KW*	gea	gea	gea	gea	—	gea	—	gea
Caloplaca erythrocarpa 13691	*Spain, Catalunya, Navarro-Rosinés s.n., BCN 13691*	EU639591	—	—	—	—	—	—	—
Caloplaca ferruginea 230	*Peru, Tarma province, Gaya 4 & Ramos, DUKE*	—	—	gea	—	—	—	—	—
Caloplaca flavescens 186	*France, Corsica, Viñas s.n., BCN*	EU639599	—	—	—	—	—	—	—
Caloplaca flavorubescens 2090/We6140		gea	gea	gea	—	—	—	—	—
Caloplaca furax 46	*Spain, Catalunya, Gaya, Fernández-Brime 557 & Llimona, BCN*	JQ301662	JQ301490	JQ301554	—	gea	gea	—	—
Caloplaca gloriae 251	*Spain, Almeria, Calatayud s.n., BCN*	EU639589	JQ301491	—	—	gea	gea	—	—
Caloplaca gloriae 59	*Spain, Almeria, P. & B. v.d. Boom 38420, P.v.d. Boom Herb.*	—	—	JQ301555	—	gea	gea	—	—
Caloplaca gr arenaria 239	*Peru, Gaya 2 & Ramos, DUKE*	gea	—	gea	—	—	—	—	—

Caloplaca gr cerina 154	*Mexico, baja California Sud, Gaya 7A & Lutzoni, DUKE*	—	—	gea	—	—	—	—	—
Caloplaca gr citrina 232	*Peru, La Merced, Gaya 26 & Ramos, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca gr citrina 240	*Peru, Tarma province, Gaya 9 & Ramos, DUKE*	—	—	gea	—	—	—	—	—
Caloplaca granulosa 13702	*Spain, Catalunya, Gaya 223 & Llimona, BCN 13702*	EU639630	gea	gea	gea	—	gea	—	gea
Caloplaca ignea 104	*USA, California, Gaya 9 & Lutzoni, DUKE*	gea	gea	gea	—	gea	gea	gea	gea
Caloplaca ignea 360	*Mexico, Baja California norte, Wetmore 72440, MIN*	EU639648	gea	—	—	—	—	—	—
Caloplaca impolita 105	*USA, California, Gaya 5 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca inconnexa 13375	*France, Roux & Navarro-Rosinés s.n, BCN 13375*	EU639647	gea	gea	—	gea	—	—	—
Caloplaca inconnexa ssp nesodes 106	*USA, California, Gaya 7 & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	—
Caloplaca isidiigera 58	*Slovakia, Šoun, CBFS*	MISSING	JQ301492	JQ301556	gea	gea	gea	—	—
Caloplaca lecanorocarpa	*Brazil, Aptroot & Caceres 4914, ISE*	KU257734	—	—	—	—	—	—	—
Caloplaca lecapsulata	*Brazil, Aptroot & Caceres 4915, ISE*	KU257735	—	—	—	—	—	—	—
Caloplaca luteoalba 155	*Mexico, Baja California Sud, Gaya 7B & Lutzoni, DUKE*	gea	gea	gea	—	—	gea	gea	gea
Caloplaca luteoalba 41	*Sweden, Gaya, Navarro-Rosinés, Arup, Søchting, Hladun & Gómez-Bolea, BCN*	gea	gea	gea	gea	—	—	—	—
Caloplaca marina 4819 331	*Finland, Stenroos 5686, H*	—	*JQ301493	*JQ301557	—	—	—	—	—
Caloplaca marina ssp marina 331	*U. Arup and M. Grube (1999).*	AF353946	—	—	—	—	—	—	—
Caloplaca maritima 243b	*Spain, Catalunya, Gaya 248b & Llimona, BCN*	EU639628	gea	gea	—	—	—	—	—
Caloplaca marmorata 416	*France, Gaya 416 & Guiedan, BCN*	EU639621	gea	gea	—	—	gea	—	—
Caloplaca micromarina	*Ukraine, Vondrák 6420, PRA*	KC611249	—	—	—	—	—	—	—
Caloplaca micromontana	*Russia, Vondrák 9467, PRA*	KC346303	—	—	—	—	—	—	—
Caloplaca microphyllina 115	*USA, California, Gaya 24 & Lutzoni, DUKE*	gea	gea	—	—	—	gea	—	—
Caloplaca microstepposa	*Ukraine, O. Nadyeina 5-14, KW*	KJ816760	—	—	—	—	—	—	—
Caloplaca microthallina 454	*Sweden, Gaya, Navarro-Rosinés, Arup, Søchting, Hladun & Gómez-Bolea, BCN*	EU639629	—	—	—	—	—	—	—
Caloplaca millegrana 200	*Antarctica, Søchting,*	—	gea	—	gea	gea	gea	gea	—
Caloplaca oasis	*NA*	gea	—	—	—	—	—	—	—
Caloplaca ochracea 21	*Spain, Catalunya, Navarro-Rosinés, BCN*	JQ301663	JQ301494	JQ301558	—	gea	—	—	—
Caloplaca orthoclada 252	*Chile, Gaya .19B & Vargas, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca paulii 183	*Spain, Catalunya, Gaya 183, BCN*	EU639606	gea	—	—	gea	gea	—	gea
Caloplaca pellodella 150	*Mexico, Baja California Sur, Gaya 2A & Lutzoni, DUKE*	—	gea	gea	gea	gea	gea	—	—
Caloplaca polycarpa 399	*France, Provença, Gaya 399, Roux & Navarro-Rosinés, BCN*	EU639631	gea	gea	gea	gea	gea	—	gea
Caloplaca proteus 13706	*Spain, Catalunya, Gaya 215 & Llimona, BCN 13706*	EU639613	gea	gea	—	gea	gea	—	—
Caloplaca pyracea 99	*USA, California, Gaya 18D & Lutzoni, DUKE*	gea	gea	gea	—	gea	gea	—	gea
Caloplaca santessoniana 242	*Chile, Gaya 37 & Vargas, DUKE*	gea	gea	gea	gea	gea	gea	gea	—
Caloplaca schistidii 245	*Spain, Catalunya, Gaya 245 & Llimona, BCN 13794*	AY233225	—	—	—	—	—	—	—
Caloplaca scopularis 374	*Finland, Stenroos 5683, H AFTOL- 4820*	—	*JQ301495	*JQ301559	—	—	—	—	—
Caloplaca scopularis 4820 374	*Scotland, Coppins 16458, O'Dare & Fryday, E 126410*	EU639646	—	—	—	gea	—	—	—
Caloplaca scoriophila 47	*Spain, Almería, P. & B. v.d. Boom 38386, P. v.d. Boom Herb.*	JQ301664	JQ301496	JQ301560	gea	gea	gea	gea	gea
Caloplaca scotoplaca 40	*SWEDEN, Bohuslän, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun, Gómez-Bolea, BCN*	JQ301665	JQ301497	JQ301561	gea	gea	gea	—	gea
Caloplaca scrobiculata 407	*Pakistan, Poelt K91-107, GZU*	EU639649	—	—	—	—	—	—	—
Caloplaca skii	*Romania, Vondrák & Šoun, CBFS*	HM582190	—	—	—	—	—	—	—
Caloplaca sol	*Wales, UK, Orange 21909, NMW*	MF595940	—	—	—	—	—	—	—
Caloplaca sp 148	*Mexico, Gaya 12 & Lutzoni, DUKE*	gea	gea	gea	—	gea	—	gea	—
Caloplaca sp 163	*Mexico, Gaya 13 & Lutzoni, DUKE*	—	gea	gea	gea	gea	gea	—	—
Caloplaca sp 167	*Mexico, Baja California, Gaya 5 & Lutzoni, DUKE*	gea	—	—	gea	—	—	—	—
Caloplaca sp 171	*Mexico, Sonora, Gaya, Lutzoni & Miranda, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 218	*Perú*	gea	—	—	gea	—	—	—	—

Caloplaca sp 220	*Peru,Tarma province, Gaya 1 & Ramos, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 221	*Peru,Tarma province, Gaya 10 & Ramos, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 225	*Peru,Tarma province, Gaya 2 & Ramos, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 228	*Peru,Tarma province, Gaya 12 & Ramos, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 234	*Peru, Lima province, Gaya 3 & Ramos, DUKE*	gea	—	—	—	—	—	—	—
Caloplaca sp 236	*Peru, Lima province, Gaya 7 & Ramos, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca sp 237	*Peru, Tarma province, Gaya 6 & Ramos, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca sp 238	*Peru, Tarma province, Gaya & Ramos, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca sp 244	*Chile, Iquique, Gaya 22B & Vargas, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca sp 245	*Chile, Gaya 35 & Vargas, DUKE*	gea	—	gea	—	—	—	—	—
Caloplaca squamosa 151	*Mexico, Baja California Sur, Gaya 2B & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	—	—
Caloplaca squamosa sl 49	*Argentina, Morris, DUKE*	JQ301666	—	JQ301562	gea	gea	gea	gea	gea
Caloplaca stillicidiorum 405	*France, Rhônes-Alpes,*	gea	gea	—	—	—	gea	—	gea
Caloplaca submergenda 107	*USA, California, Gaya 11a & Lutzoni, DUKE*	gea	gea	gea	gea	gea	—	gea	gea
Caloplaca submergenda 107	*USA, California, Gaya 11a & Lutzoni, DUKE*	gea	gea	gea	gea	gea	—	gea	gea
Caloplaca submergenda 109	*USA, California, Gaya 8a & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca submergenda 109	*USA, California, Gaya 8a & Lutzoni, DUKE*	gea	gea	gea	gea	gea	gea	gea	gea
Caloplaca syvashica	*Morocco, Soun 484, Soun*	HM582199	—	—	—	—	—	—	—
Caloplaca teicholyta 13695	*Spain, Catalunya, Gaya 203 & Navarro-Rosinés, BCN 13695*	gea	gea	—	gea	gea	gea	—	—
Caloplaca tephromelae	Australia, Tasmania G. Kantvilas 26/19 (HO—holotypus; BILAS, TUF091318—isotypi)	MW485494	MW483077	MW483076	–	–	–	–	–
Caloplaca thallincola 22	*Sweden, Halland, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun, Gómez-Bolea, BCN*	JQ301667	JQ301498	JQ301563	gea	gea	gea	gea	gea
Caloplaca thallincola 401	*Denmark, Bornholm, Søchting 9773, BCN, dupl.*	EU639604	gea	gea	gea	gea	gea	gea	—
Caloplaca thracopontica 56	*Turkey, Black Sea Coast, Šoun, CBFS*	JQ301668	JQ301499	—	—	—	gea	—	—
Caloplaca trachyphylla 385	*Pakistan, North Western Himalaya, Poelt s.n., GZU*	EU639651	gea	—	gea	gea	gea	gea	—
Caloplaca variabilis 189	*Italy, Veneto, Arup L07196, LD*	gea	gea	gea	gea	gea	—	gea	gea
Caloplaca velana 192	*Italy, Veneto, Arup L07123, LD*	gea	gea	—	gea	—	gea	—	—
Caloplaca velana 409	*France, Provença, Gaya 409 & Gueidan, BCN*	EU639609	gea	—	—	—	—	—	—
Caloplaca verruculifera 23	*Iceland, Eyjafjardarsysla, Lutzoni & Gueidan 7, DUKE*	JQ301669	JQ301500	JQ301564	gea	gea	—	gea	gea
Caloplaca xantholyta 32	*Greece, Crete, Spribille 12970, B*	JQ301670	JQ301501	JQ301565	gea	—	gea	—	—
Caloplaca zeorina	*China, Qinghai province, Oh, Han, & Hur CH140080*	KU199677	—	—	—	—	—	—	—
Caloplaca' brebissonii	Bolivia, Flakus 20411, KRAM, dupl. in LPB.	–	–	MT754212	–	–	–	–	–
Caloplaca' crocina	Bolivia, Wilk 2507, KRAM.	MT754239	MT835245	MT754207	–	–	–	–	–
Caloplaca' quadrilocularis	Bolivia, Flakus 28358, KRAM, dupl. in LPB.	MT754244	–	MT754214	–	–	–	–	–
Caloplaca' sp. 1	Bolivia, Flakus 26376 & P. Rodriguez, KRAM.	MT754242	–	MT754210	–	–	–	–	–
Caloplaca' sp. 2	Bolivia, Flakus 27643 & P. Rodriguez, KRAM.	MT754243	–	MT754213	–	–	–	–	–
Catenarina desolata	*Chile, Søchting 10436A, C*	KF657317	KF657319	—	—	—	—	—	—
Catenarina vivasiana	*Chile, Søchting 10410, C*	KF657312	KF657318	—	—	—	—	—	—
Cerothallia yorkensis	*Australia, VIC, Kärnefelt 996101, LD*	KC179101	KC179513	KC179178	—	—	—	—	—
Charcotiana antarctica	*Antarctica, Northern Victoria Land, Seppelt 25454, HO*	KJ789968	KJ789976	KJ789973	—	—	—	—	—
Cinnabaria boliviana sp. nov.	Bolivia, Wilk 3221a, KRAM holotype, LPB isotype	MT754245	MT741643	MT754215	–	–	–	–	–
Coppinsiella orbicularis	*Hungary, Vondrák, PRA*	KU554430		KU554437	—	—	—	—	—
Coppinsiella orbicularis	*Czech Rep., Vondrák, PRA*		KU554436		—	—	—	—	—
Coppinsiella substerilis	*Bulgaria, Vondrák, CBFS*	KC416108	—	—	—	—	—	—	—
Coppinsiella ulcerosa	*USA, Arup 1017, LD*	GU080298			—	—	—	—	—
Coppinsiella ulcerosa	*Tunisia, Vondrák, PRA*		KU554434	KU554439	—	—	—	—	—
Dijigiella kaernefeltiana	*Australia, Kärnefelt 20042002, LD 1238056, holotype*	KY614397		KY614445	—	—	—	—	—
Dijigiella subaggregata	*Australia, Victoria, Kärnefelt 996401, LD 1275974, holotype*	KY614398		KY614446	—	—	—	—	—
Dufourea bonae-spei	*South Africa, Feuerer & Thell 60485ab, LD*	KC179353			—	—	—	—	—
Dufourea bonae-spei	*South Africa, Feuerer & Thell 60493a, LD*		KC179516	KC179181	—	—	—	—	—
Eilifdahlia dahlii	*W Australia, Kärnefelt 20043101, LD 1270334*	KJ021318	KJ021279	KJ021253	—	—	—	—	—

Eilifdahlia wirthii	*Namibia, Wirth, Wirth & Erb, KW-L, holotype*	KJ021319	KJ021280	KJ021254	—	—	—	—	—
Fauriea chujaensis	*S Korea, Kondratyuk 141138, KoLRI 023698*	KX793097	KX793103	KX793100	—	—	—	—	—
Fauriea orientochinensis	*China, Kondratyuk, KoLRI 013959*	KX793095	KX793101	KX793098	—	—	—	—	—
Filsoniana kiamae	*Australia: South Australia, Kärnefelt 994101, LD 1101337*	KF264633	—	KF264667	—	—	—	—	—
Filsoniana lhasanensis	Z24, Wen 20110910-05	MK153161	MK439835	MK439831	–	–	–	–	–
Filsoniana rexfilsonii	*Australia: NSW, Kärnefelt 20045002, LD 1182497*	KF264638		KF264671	—	—	—	—	—
Filsoniana rexfilsonii	*Australia: NSW, Filson 20000, LD 111777*		KF264693		—	—	—	—	—
Flavoplaca mereschkowskiana	*Australia, WA, Kärnefelt 20041503, LD isotype*	KC179367	—	—	—	—	—	—	—
Fominiella tenerifensis	*Spain, Tenerife, Kondratyuk 20916, KW-L, holotype*	—	KY614478	KY614447	—	—	—	—	—
Franwilsia bastowii	*S Australia, Kärnefelt 994301, LD*	KJ021324	KJ021284	KJ021257	—	—	—	—	—
Franwilsia kilcundaensis	*W Australia, Kondratyuk 20423, LD 1237996*	KJ021326	KJ021286	KJ021259	—	—	—	—	—
Franwilsia renatae	*Namibia, Wirth & Wirth, KW-L*	KJ021329	KJ021289	—	—	—	—	—	—
Fulgensia australis BCN	*Spain, Pyrenees*	gea	gea	—	—	—	—	—	—
Fulgensia bracteata 4844	*Estonia, Saaremaa, Alanko 121555, H*	—	*JQ301502	*JQ301566	—	—	—	—	—
Fulgensia bracteata ssp bracteata AF277666	*K141, Kasalicky, Doering, Rambold & Wedin*	AF277666	—	—	—	—	—	—	—
Fulgensia canariensis 3593	*Spain, Islas Canarias, La Palma, Pérez de Paz & Hernández-Padrón s.n., TFC Lich: 3593, duplic.*	EU639587	gea	gea	gea	gea	gea	gea	gea
Fulgensia desertorum 395	*Spain, Catalunya, Gaya 395, Torra, Pérez & Galvany, BCN*	EU639585	gea	—	—	—	—	gea	gea
Fulgensia fulgens 39	*Sweden, Öland, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun, Gómez-Bolea, BCN*	JQ301671	JQ301503	JQ301567	gea	gea	gea	gea	gea
Fulgensia klementii 7480 254	*Turkey, Konya province, Kocakaya and Halici*	gea	gea	—	gea	gea	gea	gea	—
Fulgensia poeltii 400	*Spain, Catalunya, Gaya 400, BCN*	EU639586	gea	—	—	gea	—	gea	gea
Gallowayella coppinsii	*Sweden, Kaernefelt & Kondratyuk 20360, KW*	EU681359	EU680946	—	—	—	—	—	—
Gintarasiella aggregata	*South Australia, Kantvilas 476/12, HO 567239 – holotype*	KY614390	KY614479	KY614448	—	—	—	—	—
Gondwania cribrosa	*Australia, TAS, Søchting 11581, C*	KC179102	KC179526	KC179192	—	—	—	—	—
Gondwania regalis	*Antarctica, Antarctic peninsula, Søchting 11416, C*	KC179103			—	—	—	—	—
Gondwania regalis	*Antarctica, Antarctic peninsula, Søchting 11427, C*		KC179527	KC179193	—	—	—	—	—
Gondwania sejongensis	*Antarctic, Hur, KoLRI-004340, ANT 050943*	KJ133465	KJ133531	—	—	—	—	—	—
Gondwania sp. 18	*Antarctica, Antarctic peninsula, Søchting 11523, C*	KC179104		KC179194	—	—	—	—	—
Gondwania sp. 18	*Antarctica, Antarctic peninsula, Søchting 11423, C*		KC179528		—	—	—	—	—
Gyalolechia aurea	*Austria, Arup L97493, LD (ITS, SSU)*	KC179434	KC179530		—	—	—	—	—
Gyalolechia aurea	*Austria, 1993, Poelt & Grube, GZU (LSU)*			KC179196	—	—	—	—	—
Gyalolechia flavovirescens	*Austria, Arup L97253, LD (ITS); Russia, Søchting 8648, C (LSU, SSU)*	AF353966	KC179532	KC179198	—	—	—	—	—
Gyalolechia gomerana	*Spain, Søchting 9653, C*	KC179441		KC179200	—	—	—	—	—
Gyalolechia stantonii	*USA, California, Wetmore 73334, C*	KC179445	KC179535	KC179201	—	—	—	—	—
Haloplaca sp. 6	*Spain, Fuerteventura, Arup L07177, LD*	KC179295	KC179537	KC179203	—	—	—	—	—
Haloplaca suaedae	*UK, Wales, 2000, Gilbert & Giavarini, UPS isotype*	HM582197	KC179538	—	—	—	—	—	—
Hanstrassia aff lenae	*Russia, Yakovchenko T-458, KoLRI*	KY614402	KY614483	—	—	—	—	—	—
Hanstrassia jaeseounhurii	*China, Park & Hur, KoLRI 41972, holotype*	KY614399	—	—	—	—	—	—	—
Harusavskia elenkinianoides	*Chile, Oh & Hur, KoLRI 14493, holotype*	KY614403	KY614484	KY614451	—	—	—	—	—
Hosseusiella gallowayiana	*Chile, Oh, Hur, KoLRI 020689*	MG811846	—	—	—	—	—	—	—
Huea cerussata 196	*Antarctic endemic*	—	KT291519	—	KT291601	KT291658	KT291658	—	—
Huea coralligera 195	*Antarctic endemic*	gea	—	—	gea	gea	—	—	gea
Huneckia crocina	Argentina, Ferraro et al. 10823, C	MT967384	MT952902	MT952928	–	–	–	–	–
Huneckia pollinii	*U.S.A., Kansas. Morse 14464, LD 1276205*	KJ021336	KJ021296	KJ021265	—	—	—	—	—
Huneckia wrightii	Galapagos, Aptroot 63246, CDS 29981 (nrITS); Miranda 962, CDS 45159 (nrLSU, mrSSU)	MT967388	MT952903	MT952929	–	–	–	–	–
Huriella flakusii	Peru, Flakus 9263 & Cykowska (KRAM)	MN108089	MN108090	MN108238	–	–	–	–	–
Huriella loekoesiana	*South Korea, Jayalal 120433, KoLRI 15423*	KY614406	—	—	—	—	—	—	—
Ikaeria aurantiellina	*Spain, Tenerife, Kondratyuk 20916, KW-L*	KY614414	KY614493	—	—	—	—	—	—
Ikaeria serusiauxii	Macaronesia, Sipman & Aptroot B	MN586958	–	–	–	–	–	–	–
Ioplaca pindarensis 199	*China, Yunnan Prov., Sochting 9825*	gea	—	—	—	—	—	—	—
Ioplaca pindarensis 37	*China, Yunnan Prov., Aptroot 56827*	394998697	—	—	394998758	394998848	—	—	gea
Jackelixia elixii	*Australia, Kaernefelt 992002, LD*	EU681308	EU680877	—	—	—	—	—	—
Jackelixia filsonii	*Australia, Kaernefelt 20042101, LD*	EU681304	EU680881	—	—	—	—	—	—

Jackelixia streimannii	*Australia, Kondratyuk 20454, KW*	EU681306	EU680875	—	—	—	—	—	—
Jasonhuria bogilana	*South Korea, Jayalal, Park & Ryu, 120454, KoLRI 015444*	KT220196	KT220214	KT220205	—	—	—	—	—
Josefpoeltia parva 184	*Argentina, Región Mendoza, Arup L01252*	—	gea	—	—	—	—	—	—
Josefpoeltia parva 447t2	*Argentina, Honegger 447t2 (Z-ZT)*	222353465	—	—	—	—	—	—	—
Josefpoeltia sorediosa	*Chile, Frödén 1593, LD*	KC179297	KC179540	KC179205	—	—	—	—	—
Kaernefia albocrenulata	*South Africa, Wirth 41414, KW-L*	KF264648	KF264699	KF264676	—	—	—	—	—
Kaernefia gilfillaniorum	*Australia: South Australia, Kantvilas 417/11, KW-L*	KF264650	KF264701	KF264678	—	—	—	—	—
Kuettlingeria albolutescens	Italy, Toscana, J. Vondrák 8849 (PRA)	MN305804	–	MN305824	MN311941	MN311948	–	MN311933	–
Kuettlingeria areolata	Italy, Sicily, J. Vondrák, 10843 (PRA)	MN305805	MN305847	MN305825	MN311942	MN311949	–	MN311934	–
Kuettlingeria atroflava	Greece, Meteora, J. Vondrák, 8723 (PRA)	MH104921	MH100744	MH100775	MH141559	–	–	MH119858	–
Kuettlingeria cretense	Portugal, France, Haute, I. Frolov s.n.	MH104925	MH100751	MH100783	MH141560	MH119821	–	MH119845	–
Kuettlingeria diphyodes	Vienne, topotype, I. Frolov 1430	MH104926	MH100753	MH100785	MH141561	–	–	MH119846	–
Kuettlingeria emilii	Czech Republic, Pavlovske vrchy hills, J. Vondrák 9358 (PRA)	KC416102	MH100754	MH100787	MH141575	MH119822	–	MH119847	–
Kuettlingeria erythrocarpa	Russia, Black Sea coast, J. Vondrák 6650 (PRA)	MN305806	–	MN305826	MN311943	MN311950	–	MN311935	–
Kuettlingeria furax	Greece, Peloponnese, J. Vondrák 8714 (PRA)	–	–	MN305827	MN311944	–	–	MN311936	–
Kuettlingeria fuscoatroides	Italy, Sardinia, J. Vondrák 10805 (PRA)	–	MN305848	MN305828	MN311945	–	–	MN311937	–
Kuettlingeria neotaurica	Ukraine, Black Sea coast, J. Vondrák 7213 (PRA)	MN305807	MN305849	MN305829	MN311946	MN311951	–	MN311938	–
Kuettlingeria percrocata	Italy, Dolomites, J. Vondrák 4634 (PRA)	MH104931	MH100763	MH100794	MH141563	MH119823	–	MH119848	–
Kuettlingeria soralifera	France, Maritime Alps, J. Vondrák 10813	MN305808	MN305850	MN305830	MN311947	MN311952	–	MN311939	–
Kuettlingeria sp.2	Abkhazia, NP Ritsinsky, I. Frolov 1456	MH104932	MH100764	MH100786	MH141562	–	–	MH119859	–
Kuettlingeria xerica	Russia, Dagestan, 14544 (PRA)	MN305809	MN305851	MN305831	–	MN311953	–	MN311940	–
Lacrima aphanotripta	Galapagos, Aptroot 65085, CDS 31667	MT967391	–	–	–	–	–	–	–
Lacrima epiphora	Panama, van den Boom 43698, priv. herb.	MT967392	MT952904	MT952930	–	–	–	–	–
Lacrima galapagoensis	Galapagos, Bungartz 4091, CDS 28055 (nrITS); Ertz 22855, BR (nrLSU, mrSSU)	MT967397	MT952905	MT952931	–	–	–	–	–
Lacrima sonorae	Mexico, Wetmore 79435, LD	MT967405	–	–	–	–	–	–	–
Lacrima sp.	Argentina, Catamarca. Juan Martin Hernandez	MT967406	MT952906	MT952932	–	–	–	–	–
Langeottia ottolangei	*Namibia, Wirth 7660, Wirth, Erb, STU*	KJ133468	KJ133533	KJ133497	—	—	—	—	—
Lazarenkoella polycarpoides	*Iran, Haji Moniri & Parsa, KW-L*	KT456215	—	KT456230	—	—	—	—	—
Lazarenkoella zoroasteriorum	*Iran, Haji Moniri, KW-L*	KT456217	—	KT456232	—	—	—	—	—
Lazarenkoiopsis ussuriensis	*Russia, Kondratyuk, 213R3, KoLRI*	KY614419	KY614499	KY614456	—	—	—	—	—
Lecania sulphureofusca 30	*Ecuador, Zamora-Chinchipe, Sipman 53087 (B)*	—	gea	gea	—	gea	—	—	—
Lendemeriella aureopruinosa	Yakutia, Russia, holotype	MN814228.1	MW227504	MW227332	–	–	–	–	–
Lendemeriella borealis	Chelyabinsk Region, Russia, Frolov 2476	MW227317	MW227512	MW227334	–	–	–	–	–
Letrouitia aff domingensis 1	*788. 46. 60-62. A. Aptroot*	NEW	NEW	—	—	—	—	—	—
Letrouitia aff leprolyta 10	*190. 1788-89. 59. on Diospyros A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff leprolyta 107	*493. Udawalawa A. Aptroot*	—	NEW	—	—	—	—	—	—
Letrouitia aff leprolyta 11	*1467. Uduwalawa A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff leprolyta 12	*180. 2112-13. 69 A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff subvulpina 7	*149. 8312-16. 16 A. Aptroot*	NEW	NEW	—	—	—	—	—	—
Letrouitia aff subvulpina 8	*4. 8225-26. 16 A. Aptroot*	NEW	NEW	—	—	—	—	—	—
Letrouitia aff subvulpina 9	*1305. 1849-52. 61 A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff transgressa 3	*1470. 6602-07. 6613-15. Udawalawa. Sri Lanka A. Aptroot*	NEW	—	—	—	—	—	—	—
Letrouitia aff transgressa 5	*223. 7915-20. 8 A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff transgressa 6	*352. 19 A. Aptroot*	NEW	NEW	NEW	—	—	—	—	—
Letrouitia aff vulpina 2	*934. 8746-47. 53 A. Aptroot*	NEW	NEW	—	—	—	—	—	—
Letrouitia domingensis 1 55	*Dominican Republic, Duarte prov., P. & B. v.d. Boom 39385 (P.v.d. Boom Herb.)*	394998698	394998530	394998594	394998760	394998850	394998850	—	gea
Letrouitia flavidula 35	*Costa Rica, Puntarenas, Lücking 15128 (F)*	JQ301674	JQ301506	—	—	—	—	—	gea
Letrouitia parabola 11	*USA, Alabama, Gaya 11 (DUKE)*	JQ301675	JQ301507	JQ301570	JQ301720	—	JQ301765	—	gea
Letrouitia sp 206	*Peru, Gaya & Ramos 07.03.10-21 (DUKE)*	—	gea	—	gea	gea	—	—	—
Letrouitia subvulpina 44	*Costa Rica, Guanacaste, Sipman 51037e (B)*	JQ301676	—	—	—	—	—	—	gea
Letrouitia transgressa 29	*Costa Rica, Puntarenas, Sipman 51152a (B)*	—	JQ301508	—	—	—	—	—	—
Letrouitia vulpina 72	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 40279, D. & M. Brand & Sérusiaux (P.v.d. Boom Herb.)*	JQ301677	JQ301509	JQ301571	JQ301721	—	JQ301766	—	—

Loekoesia austrocoreana	*South Korea, Jayalal, Park, & Ryu 120511, KoLRI 015502*	KT220200	KT220218	KT220209	—	—	—	—	—
Massjukiella alaskana	*Canada, Talbot 1004, CAN*	EU681325	EU680911	—	—	—	—	—	—
Massjukiella kaernefeltii	*Canada, Talbot, CAN*	EU681326	EU680909	—	—	—	—	—	—
Massjukiella ucrainica	*Denmark, Corfixen et al., H*	EU681327	EU680916	—	—	—	—	—	—
Megaloblastenia marginiflexa 69	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 39805, D. & M. Brand & Sérusiaux (P.v.d. Boom Herb.)*	JQ301678	JQ301510	JQ301572	JQ301722	—	JQ301767	gea	—
Megalospora admixta 42	*Ecuador, Zamora-Chinchipe, Sipman 52536 (B)*	—	JQ301511	—	—	—	—	gea	—
Megalospora coccodes 25	*Thailand, Chiang Mai, Sipman 48671 (B)*	JQ301679	JQ301512	—	—	—	—	—	—
Megalospora coccodes 94	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 40221, D. & M. Brand & Sérusiaux (P.v.d. Boom Herb.)*	gea	gea	gea	—	—	—	—	—
Megalospora isidiata 93	*Mascarene Islands, La Réunion, P. & B. v.d. Boom 40469, D. & M. Brand & Sérusiaux (P.v.d. Boom Herb.)*	gea	gea	—	—	—	—	gea	—
Megalospora pruinata 43	*Indonesia, West Java, Sudirman & Sipman 51496 (B)*	JQ301680	JQ301513	—	—	—	—	gea	—
Megalospora sulphurata 73	*Mascarene Islands, La Réunion, P. v.d. Boom 40349, M. Brand & Sérusiaux (P.v.d. Boom Herb.)*	JQ301681	JQ301514	JQ301573	—	—	JQ301768	gea	gea
Megalospora tuberculosa 107	*Costa Rica , Puntarenas, Sipman 47933 (DUKE)*	336397114	46411388	46411439	113707385	113957833	113957833	—	—
Megalospora tuberculosagr 1 52	*Dominican Republic, Puerto Plata prov., P. & B. v.d. Boom 39047 (P.v.d. Boom Herb.)*	JQ301682	JQ301515	JQ301574	JQ301723	—	JQ301769	gea	—
Mikhtomia gordejevii	*Republic of Korea, Hur 080515, KoLRI 008744*	KJ021231	KJ021307	—	—	—	—	—	—
Mikhtomia multicolor	*SKA14, Republic of Korea, Kondratyuk 121384, KoLRI 016442*	KJ021237	—	KJ021272	—	—	—	—	—
Mikhtomia yeosuensis	*Republic of Korea, Jayalal 120360, KoLRI 015350*	KJ021235	—	KJ021270	—	—	—	—	—
Nevilleiella lateritia	*Australia, Kärnefelt 990904, LD 1236736*	KY614427	KY614502	KY614464	—	—	—	—	—
Nevilleiella marchantii	*Western Australia, Kondratyuk 20444, KW-L – isotype*	KY614425	KY614500	KY614462	—	—	—	—	—
Oceanoplaca caesioisidiata	Cape Verde, van den Boom 36538, LD	MT967411	MT952910	–	–	–	–	–	–
Oceanoplaca caesiosorediata	Cape Verde, van den Boom 36346, LD	MT967412	MT952911	MT952936	–	–	–	–	–
Oceanoplaca catillarioides	Cape Verde, van den Boom 36365, LD	MT967410	MT952909	MT952935	–	–	–	–	–
Oceanoplaca chemoisidiosa	Galapagos, Bungartz 6436, CDS 34651 (nrITS, nrLSU); Aptroot 64354, CDS 30919 (mrSSU)	MT967413	MT952912	MT952937	–	–	–	–	–
Oceanoplaca isidiosa	Galapagos, Adsersen, LAM8-1, C	MT967416	MT952913	MT952938	–	–	–	–	–
Oceanoplaca sideritoides	Galapagos, Bungartz 3663, CDS 27481	MT967424	MT952915	MT952939	–	–	–	–	–
Oceanoplaca sp.	Panama, van den Boom 44028, LD	MT967429	MT952917	MT952941	–	–	–	–	–
Olegblumia demissa	*Italy, Arup L97911, LD*	AF353960	KC179505	KC179172	—	—	—	—	—
Opeltia juniperina	*China, Oh & Hur, CH 130112 KoLRI*	KY614429	KY614504	—	—	—	—	—	—
Opeltia neobaltistanica	*China, Oh & Hur, CH 130098 KoLR*	KY614428	KY614503	—	—	—	—	—	—
Orientophila corticola	South Korea, B.G. Lee, 2019-000118 (holotype, BDNA-L-0000118)	MN687909.1	–	–	–	–	–	–	–
Orientophila infirma	Yakutia, Russia, holotype	MN814235	MW227507	MW227329	–	–	–	–	–
Orientophila sp. 21	*Japan, Frisch Jp171, LD*	KC179372	KC179544	KC179210	—	—	—	—	—
Orientophila sp. 22	*Japan, Thor 25400, UPS*	KC179373	KC179545	KC179211	—	—	—	—	—
Ovealmbornia reginae	*South Africa, Kondratyuk, KW-L*	KT805130	KT805134	KT805132	—	—	—	—	—
Ovealmbornia volk-marwirthii	*South Africa, Wirth, STU*	KJ133477	KJ133541	KJ133501	—	—	—	—	—
Oxneria alfredii	*Russia, Kondratyuk, KW*	EU681344	EU680933	—	—	—	—	—	—
Oxneria sp.	*Russia, Urbanavichus, H*	EU681343	—	—	—	—	—	—	—
Oxneriopsis aff oxneri	*Russia, Kondratyuk 21138, KW-L*	KY614431	KY614506	—	—	—	—	—	—
Oxneriopsis oxneri	*Russia, Kondratyuk 21138, KW-L*	KY614430	KY614505	—	—	—	—	—	—
Pachypeltis intrudens	*Afghanistan, Soelberg s.n., C*	KC179107	—	—	—	—	—	—	—
Pachypeltis invadens	*Norway, Svalbard, Elvebakk 03:109, TROM*	KC179108	KC179548	KC179212	—	—	—	—	—
Pachypeltis sp. 23	*China, Abbas & Xahidin 500002, XJUG*	KC179109	KC179549	KC179213	—	—	—	—	—
Pachypeltis sp. 24	*China, Abbas & Mahamat 500113, XJUG*	KC179110	KC179550	KC179214	—	—	—	—	—
Parvoplaca athallina	*Antarctica, Søchting 11393, C*	KC179111	—	—	—	—	—	—	—
Parvoplaca nigroblastidia	*Turkey, Vondrák JV10617, PRA*	KT161978	—	—	—	—	—	—	—
Parvoplaca servitiana	*Greece, Spribille 16225, PRA*	JN641778	—	—	—	—	—	—	—
Parvoplaca sp. 26	*Sweden, Arup L10208, LD*	KC179113	KC179551	KC179215	—	—	—	—	—
Parvoplaca suspiciosa	*Russia, Hermansson 16839, LD*	KC179115	—	—	—	—	—	—	—
Parvoplaca tiroliensis	*Sweden, Arup L02364, LD (ITS); Sweden, Frödén 1945, LD (LSU, SSU)*	KC179116	KC179552	KC179216	—	—	—	—	—
Phaeoplaca camptidia	USA, Kansas, Morse 14420, LD	MT967431	MT952918	MT952942	–	–	–	–	–
Phaeoplaca tortuca	Galapagos, Aptroot 65189, CDS 31773	MT967433	–	MT952943	–	–	–	–	–

Pyrenodesmia albopruinosa	Italy, Verona, TSB 37658 J. Vondrák	EF093577	–	MH100770	MH141578	MH119824	–	MH119851	–
Pyrenodesmia albopustulata	Turkey, Black Sea region, 10463 (PRA)	MH104918	MH100741	MH100771	MH141564	MH119825	–	–	–
Pyrenodesmia alociza	Italy, Ascoli Piceno, TSB 37735	EF090931	MH100742	MH100772	MH141587	MH119826	–	MH119860	–
Pyrenodesmia alociza	*Slovenia, Wilfling 3471 and Arup, GZU*	AY313973	—	—	—	—	—	—	—
Pyrenodesmia atroalba	USA, Montana, T. Spribille s.n.	MH104920	MH100743	MH100774	MH141565	MH119827	–	MH119861	–
Pyrenodesmia badioreagens	Italy, Foggia, TSB 36422	EF081035	MH100745	MH100776	MH141566	MH119828	–	MH119862	–
Pyrenodesmia circumalbata	Turkey, Mersin, M.G. Halici s.n.	MH104923	MH100748	MH100780	MH141567	MH119831	–	MH119865	–
Pyrenodesmia concreticola	Kazakhstan, Mangistau, J. Vondrák 9443 (PRA)	KC884506	MH100749	MH100781	MH141568	MH119832	–	MH119852	–
Pyrenodesmia erodens	Turkey, Kahramanmaraş, 12733 (PRA)	MH104927	MH100755	MH100788	MH141569	MH119833	–	MH119853	–
Pyrenodesmia erodens	*Italy, Tretiach & Pinna, TSB 34131*	AY313978	—	—	—	—	—	—	—
Pyrenodesmia helygeoides	Switzerland, Ticino, I. Frolov 1414	MH104929	MH100757	MH100790	MH141570	MH119834	–	MH119866	–
Pyrenodesmia micromarina	Ukraine, Black Sea coast, J. Vondrák 7236 (PRA)	KC611248	MH100758	MH100791	–	MH119835	–	MH119867	–
Pyrenodesmia micromontana	Russia, Southern Ural, holotype, J. Vondrák 9467 (PRA)	KC346303	MH100759	MH100792	MH141580	MH119836	–	MH119868	–
Pyrenodesmia microstepposa	Czech Republic, Bohemian karst, holotype, J. Vondrák 9141 (PRA)	KC984530	MH100760	–	MH141581	MH119837	–	MH119869	–
Pyrenodesmia molariformis	Ukraine, Luhansk, O. Nadyeina 132 (KW)	KC416143	MH100761	MH100793	MH141582	MH119838	–	–	–
Pyrenodesmia peliophylla	USA, California, K. Knudsen 13557 (UCR)	MH104930	–	–	MH141571	MH119839	–	MH119870	–
Pyrenodesmia pratensis	USA, Wyoming, in MIN 891605	MH104933	MH100765	MH100795	MH141583	MH119840	–	MH119871	–
Pyrenodesmia sp. 1	Czech Republic, Bohemian karst, J. Vondrák 9673 (PRA)	KC884525	-	MH100778	MH141579	MH119829	–	MH119863	–
Pyrenodesmia tianshanensis	China, Xinjiang, holotype, XJU 1691	GU552277	–	MH100798	MH141585	MH119841	–	MH119872	–
Pyrenodesmia transcaspica	Kazakhstan, Mangistau, J. Vondrák 9430 (PRA)	MH104936	MH100768	MH100799	MH141572	MH119842	–	MH119873	–
Raesaeneniana maulensis	*Chile, Wang, Oh & Hur, KoLRI- 014500, CL 120331*	—	KT456248	—	—	—	—	—	—
Rehmanniella wirthii	*Namibia, Wirth & Heklau, STU*	MG811851	MG811853	MG811852	—	—	—	—	—
Rufoplaca tristiuscula	*Norway, Arup L08171, LD*	KC179460	KC179575	KC179237	—	—	—	—	—
Sanguineodiscus bicolor	Russia, Altai, J. Vondrák 10373 (PRA)	MH104922	MH100746	MH100777	MH141577	–	–	MH119850	–
Sanguineodiscus haematites	Ukraine, Black Sea coast, J. Vondrák 7278 (PRA)	MH104928	MH100756	MH100789	MH141558	MH119820	–	MH119856	–
Sanguineodiscus viridirufus	Czech Republic, Hanušovice, J. Vondrák 6702 (PRA)	MH104919	–	MH100773	MH141574	–	–	MH119857	–
Scutaria andina	Bolivia, Wilk 3178, KRAM, dupl. in LPB	MT754221	MT741623	MT754190	–	–	–	–	–
Scutaria andina	*Argentina, Frödén 1669, LD*	KC179298	KC179581	KC179242	—	—	—	—	—
Seawardiella tasmaniensis	*Australia, Tasmania, Karnefelt 9977001, LD*	KJ133448	KJ133506	KJ133487	—	—	—	—	—
Seirophora blumii	*Iran, Haji Moniri, KW-L*	KT456219	KT456249	KT456234	—	—	—	—	—
Seirophora californica 153	*Mexico, Gaya 03.04.10-9 & Lutzoni (DUKE)*	gea	gea	gea	gea	gea	—	gea	gea
Seirophora californica 80	*USA, California, Bratt 11947, Lichenes Exsiccati - Nash #394*	gea	gea	—	—	gea	gea	—	—
Seirophora contortuplicata 197	*Tadjikistan, 4000m, 14.09.2010, leg. U. Sochting*	gea	gea	—	gea	gea	—	—	—
Seirophora contortuplicata 27	*Spain, Catalonia, Navarro-Rosinés & Gaya 199 (BCN 13686)*	gea	—	—	—	—	—	—	—
Seirophora contortuplicatus 27 EU639614	*Iran, Khorassan prov., Maassoumi, Sohrabi & Safavi 2275 (B)*	—	JQ301516	—	—	—	JQ301770	—	—
Seirophora lacunosa 10	*Spain, Alacant, Pérez s.n. (BCN, herb. Gaya 391)*	JQ301683	JQ301517	JQ301575	JQ301724	JQ301771	JQ301771	—	—
Seirophora scorigena	*Spain, Lanzarote, S. and B Snogerup 17201, LD Cap*	KC179466	KC179583	KC179244	—	—	—	—	—
Seirophora villosa	*Morocco, Guttová, KW-L ex BRA*	KY614436	KY614511	KY614469	—	—	—	—	—
Shackletonia buelliae	*Antarctica, South Shetland Isl., Søchting 7583, C*	KC179117	KC179578	—	—	—	—	—	—
Shackletonia cryodesertorum	*Antarctica, McMurdo Dry Valleys, De los Rios 18379, MA-Lich*	JX036143	KU599932	KU599931	—	—	—	—	—
Shackletonia hertelii	*Chile, Søchting 10349, C*	KC179118	KC179579	KC179240	—	—	—	—	—
Shackletonia sauronii	*Antarctica, South Shetland Isl., Søchting 7654, C*	KC179120	KC179580	KC179241	—	—	—	—	—
Sirenophila bermaguiana	*Australia, NSW, Kondratyuk 20487, LD isotype*	KC179299	KC179584	KC179245	—	—	—	—	—
Sirenophila cliffwetmorei	*Australia, Kantvilas 446/12 KW-L ex HO*	KY614438	KY614513	KY614471	—	—	—	—	—
Sirenophila gallowayii	*Australia, Tasmania, Kärnefelt 997504, LD isotype*	KC179301	KC179586	KC179247	—	—	—	—	—
Sirenophila gintarasii	*Australia, Kondratyuk, KW-L*	KY614437	KY614512	KY614470	—	—	—	—	—
Sirenophila jackelixii	*Gea Zealand, Kärnefelt 998604, LD*	KC179303	KC179587	KC179248	—	—	—	—	—
Sirenophila ovis-atra	*Chile, Zeuthen Søgaard 90, C*	KC179306	KC179589	KC179250	—	—	—	—	—
Solitaria chrysophthalma	*Sweden, Arup L03101, LD*	KC179408	KC179590	KC179251	—	—	—	—	—
Squamulea chelonia	Galapagos, Bungartz 4521, CDS 28607	MT967448	–	–	–	–	–	–	–
Squamulea flakusii	Galapagos, Bungartz 4131, CDS 28162	MT967442	–	–	–	–	–	–	–
Squamulea humboldtiana	Galapagos, Bungartz 5151, CDS 29364	MT967437	–	–	–	–	–	–	–
Squamulea oceanica	Galapagos, Yánez-Ayabaca 2023, CDS 48373 (holotype)	MT967445	–	–	–	–	–	–	–
Squamulea phyllidizans	Galapagos, Aptroot 65468, CDS 32057	MT967456	–	–	–	–	–	–	–
Sucioplaca diplacia	Galapagos, Ertz 11610, CDS 36936	MT967470	MT952922	MT952945	–	–	–	–	–

Tarasginia whinrayi	*South Australia: Kangaroo Island, Kantvilas 464/12, de Villiers, KW-L-ex HO 567227*	KT456221	—	KT456236	—	—	—	—	—
Tassiloa digitaurea	*Chile, Patagonia, Oh, Hur CL 130214, KoLRI 017647*	KP096222	KP096224	—	—	—	—	—	—
Tassiloa magellanica	Chile, Søchting 12265, C	MW397534	MW397530	MW397526	–	–	–	–	–
Tayloriella erythrosticta	*Western Australia, Kondratyuk 20435, LD*	KT456222	—	KT456237	—	—	—	—	—
Teloschistes cf chrysophtalmus 147	*USA, Mexico, Gaya 03.04.10-7 & Lutzoni (DUKE)*	gea	gea	gea	gea	—	—	—	—
Teloschistes cf chrysophthalmus 248	*Chile, Chiloé province, Gaya 27.02.11-1, R. Vargas (DUKE)*	gea	—	gea	—	—	—	—	—
Teloschistes cf nodulifer 216	*Peru, Tarma province, Gaya 07.05.10-7, D. Ramos (DUKE)*	—	gea	gea	gea	gea	gea	—	gea
Teloschistes chrysophtalmus 06 1	*Spain, Gaya 200b & X. Llimona (BCN 13687)*	EU639652	JQ301518	JQ301576	JQ301725	JQ301772	JQ301772	gea	—
Teloschistes exilis 12	*USA, Texas, Hillis 07-726 (DUKE)*	JQ301684	JQ301519	JQ301577	JQ301726	JQ301773	JQ301773	gea	—
Teloschistes fasciculatus 427	*Australia, Victoria, Elix 40388 & Streimann (DUKE 297)*	EU639653	gea	—	—	—	—	—	—
Teloschistes flavicans 102	*USA, California, Gaya 02.26.10-8, Lutzoni (DUKE)*	—	gea	gea	gea	gea	gea	gea	gea
Teloschistes flavicans 103	*USA, California, Gaya 02.26.10-8, F. Lutzoni (DUKE)*	gea	gea	gea	gea	—	gea	—	—
Teloschistes hosseusianus 50	*Argentina, Villaciencio, Morris (DUKE)*	JQ301686	JQ301521	JQ301579	JQ301728	—	JQ301775	gea	gea
Teloschistes hypoglaucus	*Kenya, Frisch 94/38a, Frisch, priv. herb.*	KC179319	KC179595	—	—	—	—	—	—
Teloschistes siberianus 423	*Australia, Gea South Wales, Lumbsch & Streimann s.n. (DUKE, A. Vezda: Lichenes rariores exsiccati 30)*	EU639655	gea	—	—	gea	—	—	—
Teloschistopsis bonae-spei	*South Africa, Sipman 20289, B*	KC179322	KC179596	KC179257	—	—	—	—	—
Teloschistopsis eudoxa	*Namibia, Crittenden s.n., LD*	KC179324	KC179597	KC179258	—	—	—	—	—
Teuvoahtiana fernandeziana	*Chile, Wang, Oh & Hur, CL-120121*	KY614440	KY614515	KY614474	—	—	—	—	—
Teuvoahtiana rugulosa	*Chile, 30.01.2012, Wang, X. Y., Oh, S.-O., Hur, J.-S., CL-120331 KoLRI 14500*	KY614443	KY614518	—	—	—	—	—	—
Thelliana pseudokiamae	*South Australia, Kärnefelt 994101, LD 1101337*	KT456225	—	KT456240	—	—	—	—	—
Tomnashia luteominia	*USA, California, Wetmore 73797, LD*	KC179387			—	—	—	—	—
Tomnashia luteominia	*USA, California, Søchting 11219, C*		KC179556	KC179220	—	—	—	—	—
Tomnashia rosei	*USA, California, Arup L89165, LD (ITS)*	KC179390			—	—	—	—	—
Tomnashia rosei	*USA, California, Søchting 11225, C (LSU, SSU)*		KC179559	KC179223	—	—	—	—	—
Upretia amarkantakana	*India, Kondratyuk, LWG*	MG652764	MG652767	—	—	—	—	—	—
Upretia squamulosa	*China, Yunnan Prov. Wang Lisong et al. 16-50148 (KUN)*	MH497056	—	MH497052	MH472573	MH472575	—	—	—
Villophora darwiniana	Chile, Søchting 10439, C	MW397537	MW397533	MW397533	–	–	–	–	–
Villophora isidioclada	*Chile, Søchting 10185, C*	KC179325	KC179606	KC179266	—	—	—	—	—
Villophora onas	Chile, Søchting 10441b, C	MW397554	–	–	–	–	–	–	–
Villophora patagonica	Chile, Søgaard 9, C	MW397558	–	–	–	–	–	–	–
Villophora rimicola	Antarctica, Søchting 11425, C	MW397562	–	–	–	–	–	–	–
Villophora sp. 1	Bolivia, Wilk 3206, KRAM.	MT754241	MT741642	MT754209	–	–	–	–	–
Villophora sp. 48	*Chile, Søchting 10399, C*	KC179328	KC179607	KC179267	—	—	—	—	—
Villophora sp. 50	*Chile, Frödén 1563, LD*	KC179330	—	KC179268	—	—	—	—	—
Villophora_wallaceana	Chile, Søgaard 68, C	MW397564	–	–	–	–	–	–	–
Wetmoreana sp. 1	Bolivia, Wilk 3236a, KRAM.	MT754247	MT741645	MT754217	–	–	–	–	–
Wetmoreana sp. 2	Peru, Flakus 9279 & Cykowska, KRAM.	MT754225	MT741628	MT754194	–	–	–	–	–
Wetmoreana sp. 3	Peru, Flakus 9268 & Cykowska (KRAM).	MT754231	MT741634	MT754200	–	–	–	–	–
Wetmoreana sp. 4	Peru, Flakus 9414 & Cykowska, KRAM.	MT754222	MT741624	MT754191	–	–	–	–	–
Wetmoreana sp. 52	*Chile, Frödén 1645, LD*	KC179334	KC179609	KC179270	—	—	—	—	—
Xanthodactylon flammea 95	*South Africa, Skeleton Coast, Crittenden*	gea	gea	—	—	—	—	—	—
Xanthodactylon flammeum	*South Africa, Feuerer & Thell 36/1940, HBG*	EU681315	EU680898	—	—	—	—	—	—
Xanthodactylon inflata	*South Africa, Feuerer & Thell, HBG*	EU681317	EU680899	—	—	—	—	—	—
Xanthomendoza borealis 029 3451	*Russia, Zhurbenko 94411 (UPS)*	—	—	39752532	—	—	—	—	—
Xanthomendoza borealis 77	*East Antarctica, Hallett Peninsula, Crittenden (DUKE)*	gea	gea	gea	—	gea	gea	gea	gea
Xanthomendoza borealis FNM029	*Russia,Urbanavichus (H; A. Thell s.n.)*	197342754	196174865	—	—	—	—	—	—
Xanthomendoza fallax 33	*Alaska, Gaya, Geml, Gomà (DUKE)*	JQ301687	JQ301522	JQ301580	JQ301729	JQ301776	JQ301776	—	gea
Xanthomendoza fulva 251		—	—	gea	—	—	—	—	—
Xanthomendoza fulva 85	*USA, Gea Jersey, Lendemer 613 & Moore (PH, ASU)*	gea	—	—	—	—	—	—	—
Xanthomendoza fulva 98	*USA, California, Gaya 03.01.10-18C & F. Lutzoni (DUKE)*	—	gea	gea	gea	—	gea	gea	gea
Xanthomendoza galericulata	*USA, Sharnoff & Sharnoff 1186.01, LD*	EU681357	EU680945	—	—	—	—	—	—
Xanthomendoza hasseana 5 120	*USA, Oregon*	gea	—	—	—	—	—	—	—
Xanthomendoza kashiwadanii	Argentina, Frödén xxxx, LD	MT967475	–	–	–	–	–	–	–
Xanthomendoza leoncita	Galapagos, Bungartz 4417, CDS 28502 (holotype)	MT967476	–	–	–	–	–	–	–
Xanthomendoza mendozae 174	*USA, California, Knudsen 13355.2 (UCR 2278KK64)*	—	—	gea	—	gea	—	—	—
Xanthomendoza mendozae 20	*Bolivia, Potosí, Reeb, VR 11- V-02-14 (DUKE)*	JQ301688	JQ301523	JQ301581	—	—	JQ301777	gea	gea
Xanthomendoza mendozae 203		—	gea	—	gea	—	—	—	—

Xanthomendoza montana 84	*USA, Arizona, Sweat KGS136A, Bates, Iselin & Nash (ASU 574356)*	gea	gea	gea	—	—	—	gea	gea
Xanthomendoza novozelandica 50t1	*Gea Zealand, South Island, Honegger 50t1 (Z+ZT)*	222353463	—	—	—	—	—	—	—
Xanthomendoza oregana 173	*USA, Oregon, McCune 29897*	gea	—	—	—	—	—	—	—
Xanthomendoza poeltii 7	*Sweden, Halland, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun, Gómez-Bolea (BCN)*	JQ301689	JQ301525	JQ301583	JQ301730	JQ301778	JQ301778	gea	gea
Xanthomendoza soechtingii	Russia, Søchting, US6422, C	MT967478	–	–	–	–	–	–	–
Xanthomendoza sogdiana	*Iran, Sohrabi 1343, KW*	EU681354	EU680942	—	—	—	—	—	—
Xanthomendoza sp 243		gea	—	gea	—	—	—	—	—
Xanthomendoza ulophyllodes 78	*USA, Wisconsin, Egan EL-15, 811 (OMA, ASU)*	gea	gea	gea	—	gea	gea	—	—
Xanthomendoza weberi 201	*USA, North Carolina, Perlmutter 2684 (NCU)*	gea	gea	—	gea	—	—	gea	gea
Xanthomendoza weberi 202	*USA, North Carolina, Perlmutter 2685 (NCU)*	—	gea	—	gea	—	gea	gea	—
Xanthopeltis rupicola 246	*Chile, Metropolitan Region, Vargas 3638 & Ugarte*	—	gea	gea	gea	—	gea	gea	gea
Xanthoria andina 31	*Bolivia, Dpto. La Paz, Canseco A-113 (B)*	gea	gea	gea	—	gea	gea	—	—
Xanthoria aureola 9	*SWEDEN, Bohuslän, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun & Gómez-Bolea (BCN)*	JQ301690	JQ301526	JQ301585	JQ301731	JQ301780	JQ301780	—	gea
Xanthoria calcicola 3	*Spain, Catalonia, Llimona, Fernández-Brime (BCN)*		JQ301527	JQ301586	—	JQ301781	JQ301781	gea	gea
Xanthoria calcicola 3 U257	*U257; U. Arup and M. Grube*	AF353944	—	—	—	—	—	—	—
Xanthoria candelaria 2	*Sweden, Halland, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun & Gómez- Bolea (BCN)*	—	JQ301528	JQ301587	JQ301732	JQ301782	JQ301782	gea	gea
Xanthoria candelaria 2 8444	*France, Auvergne, Llimona, Gómez-Bolea & Navarro-Rosinés (BCN 8444)*	EU639625	—	—	—	—	—	—	—
Xanthoria cf concinna 97	*USA, California, Gaya 03.01.10-18B & F. Lutzoni (DUKE)*	—	gea	gea	gea	gea	gea	gea	—
Xanthoria elegans 1 192	*Spain, Catalunya, Gaya 192 (BCN)*	EU639642	—	—	—	—	—	—	—
Xanthoria elegans 13	*USA, Alaska, Gaya, Geml, Gomà (DUKE)*	—	gea	gea	gea	gea	gea	gea	gea
Xanthoria flammea		—	—	—	—	—	—	—	—
Xanthoria karrooensis 5627 066	*South Africa, W. & U. Wetschnig (GZU)*	EU681320	EU680903	—	—	—	—	—	—
Xanthoria karrooensis AF5627		EU681321	—	—	—	—	—	—	—
Xanthoria ligulata 5628 011	*Australia, Kondratyuk 20481 (KW)*	EU681302	EU680872	—	—	—	—	—	—
Xanthoria parietina 8	*Sweden, Bohuslän, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun, Gómez-Bolea (BCN)*	JQ301691	JQ301530	JQ301589	JQ301734	JQ301784	JQ301784	—	gea
Xanthoria polycarpa 4	*Switzerland, Zurich, R. Honegger 215 (Z+ZT)*	222352275	gea	gea	gea	gea	—	gea	gea
Xanthoria polycarpa 4 Hon215	*Sweden, Öland, Gaya, Navarro-Rosinés, Arup, Sochting, Hladun & Gómez-Bolea (BCN)*	—	JQ301531	JQ301590	JQ301735	JQ301785	—	—	—
Xanthoria resendei 233b	*Spain, Islas Canarias, Hladun & Gómez-Bolea s.n. (BCN)*	EU639641	—	—	—	—	—	—	—
Xanthoria sorediata 367	*Spain, Catalunya, Gaya 367, Gómez-Bolea & Ariño (BCN)*	EU639643	gea	—	—	—	—	—	—
Xanthoria sorediata 5635 046	*Finland, Haikonen 24238 (H)*	EU681335	EU680920/	—	—	—	—	—	—
Xanthoria sp 249		—	—	gea	—	—	—	—	—
Xanthoria sp 76		gea	gea	gea	gea	gea	—	—	gea
Xanthoria tenax 96	*USA, California, Gaya 03.01.10-18A & F. Lutzoni (DUKE)*	—	gea	gea	gea	gea	gea	gea	—
Xanthoria turbinata 119	*South Africa, Feuerer & Thell (HBG)*	EU681313	EU680895	—	—	—	—	—	—
Yoshimuria galbina	*Republic of Korea, Wang & Ryu 110482, KoLRI 013527*	KJ021250		—	—	—	—	—	—
Yoshimuria galbina	*Republic of Korea, Wang & Ryu 110519, KoLRI 013563*		KJ023197	—	—	—	—	—	—
Yoshimuria spodoplaca	*Republic of Korea, Wang & Ryu 110364, KoLRI 012931*	KJ021249	KJ023194	—	—	—	—	—	—
Zeroviella coreana	*Republic of Korea, Kondratyuk 02, Lőkös, Oh & Hur 140327, KoLRI 022689*	KU056848	KU043373	—	—	—	—	—	—
Zeroviella esfahanensis	*Iran: Karkas, Zarei-Darki 1336, KW-L*	KU056849	KU043374	—	—	—	—	—	—
Zeroviella laxa	*China, Hur, KoLRI 008988*	KU056853	KU043378	—	—	—	—	—	—
Zeroviella mandschurica	*China, Hur, KoLRI 010705*	KU056855	—	—	—	—	—	—	—
Zeroviella papillifera	*Ukraine, Kondratyuk 20350 et al., KW-L*	EU681331	EU680927	—	—	—	—	—	—
Zeroviella ussurica	*Russia, Kondratyuk RU130397, KoLRI 019924*	KU056857	KU043379	—	—	—	—	—	—

Table A.5: Query genes used to pull six marker loci from assembled lichen genomes. 'gea' are unpublished sequences from Ester Gaya

Gene	**NCBI accession**	**Query taxon**
mtSSU	JQ301530.1	*Xanthoria parietina*
nucLSU	JQ301589	*Xanthoria parietina*
RPB1	JQ301725	*Teloschistes chrysophthalmus*
RPB2	JQ301784	*Xanthoria parietina*
RET1	gea	*Xanthoria parietina*
MCM7	gea	*Teloschistes chrysophthalmus*

Results

Table A.6: Genome assembly statistics for the 59 new assembled mycobiont genomes in Chapter 3.

Taxon	Assembly size (Mbp)	Scaffolds	N50 (K)	BUSCO %	Dup. BUSCO %	Genes	Accession
Blastenia crenularia	27.59	1754	35.33	90.39	0.06	8341	LIQ216CACR
Brigantiaea leucoxantha	25.00	1554	24.88	86.17	0.23	8040	LIQ191BRLE
Calogaya cf. arnoldii	37.46	6356	11.17	87.40	0.35	9372	LIQ161CABI
Calogaya cf. arnoldii	29.50	172	329.50	92.44	0.41	9108	LIQ222CASA
Caloplaca s.l. sp. 1	36.73	419	147.62	93.61	0.06	7786	LIQ188CASP
Caloplaca s.l. sp. 2	41.56	3715	23.85	89.92	0.18	8166	LIQ190CASP
Caloplaca s.l. sp. 3	32.97	311	189.87	91.44	0.06	7672	LIQ195CASP
Dirinaria leoncita	27.76	3110	14.96	83.12	0.06	8539	LIQ203DILE
Flavoplaca austrocitrina	40.51	12649	4.10	77.96	1.52	11200	LIQ221CAAUS
Flavoplaca cf. flavocitrina	27.80	1513	42.94	87.92	0.23	8178	LIQ226CATH
Flavoplaca citrina	42.23	7384	10.13	88.04	0.23	8053	LIQ218CACI
Flavoplaca holocarpa	35.18	547	113.19	84.06	0.64	9314	LIQ219CAHO
Flavoplaca limonia	27.30	3555	15.31	85.05	0.23	8017	LIQ243FLLI
Flavoplaca navasiana	41.76	318	225.09	87.34	0.82	8985	LIQ225CANA
Flavoplaca navasiana	34.96	366	153.85	80.72	0.41	9863	LIQ249FLNA
Gyalolechia fulgens	31.77	3750	13.84	87.75	0.18	7430	LIQ77FFU
Gyalolechia gomerana	32.21	1597	30.54	90.80	0.00	7649	LIQ248ELGO
Gyalolechia sp. 1	30.65	1908	24.75	86.64	0.23	7125	LIQ233GYSU
Gyalolechia sp. 2	27.36	1722	24.99	90.04	0.00	7206	LIQ234HUPO
Haloplaca sp. 1	24.65	5521	7.34	80.01	0.29	7501	LIQ224CACE
Haloplaca sp. 2	28.42	9667	3.67	76.96	0.35	8037	LIQ235IKAU
Heterodermia aff. obscurata	23.82	7795	3.70	91.15	0.23	9615	LIQ199HEOB
Ikaeria aurantiellina	27.01	1279	31.62	87.51	0.06	7907	LIQ238CAAE
Kuettlingeria erythrocarpa	29.31	2426	24.72	90.91	0.12	8528	LIQ228CAER
Letrouitia sp. 1	23.36	6361	4.79	75.62	0.06	8078	LIQ88LEMU
Letrouitia subvulpina	25.24	997	47.09	88.28	0.41	8059	LIQ89LESU
Loekoesia sp. 1	26.26	2071	25.77	86.34	0.06	8059	LIQ187CASP
Polycauliona sp. 1	26.91	1483	31.39	89.39	0.06	8349	LIQ167XMSP
Polycauliona sp. 2	37.06	14348	3.14	90.39	0.18	8673	LIQ246FLCI
Pyrenodesmia sp. 2	25.33	1732	79.51	87.05	0.59	8537	LIQ192CASP
Pyrenodesmia sp. 3	24.94	3152	14.81	94.72	0.18	8748	LIQ232PYVA
Rusavskia elegans	30.50	1578	32.81	87.57	0.06	8141	LIQ193RUEL
Seirophora lacunosa	24.97	645	133.42	89.68	0.12	7403	LIQ92SELA
Seirophora scorigena	47.99	2481	29.99	87.98	0.00	7286	LIQ250SESC
Seirophora villosa	23.71	4238	8.52	82.71	0.06	7302	LIQ240SEVIL
Squamulea sp. 1	37.70	744	150.68	91.97	0.18	7930	LIQ196BLSP
Squamulea squamosa	29.29	595	83.46	90.74	0.12	8373	LIQ197SQSQ
Tayloriellina microphyllina	26.62	2534	19.86	87.92	0.00	8233	LIQ184CAMI
Teloschistes cf. exilis	24.44	2171	18.61	87.34	0.29	7663	LIQ194TCHR
Teloschistes chrysophthalmus	25.93	569	97.35	96.89	0.00	8217	LIQ125TCHR2
Teloschistes exilis	24.40	1994	19.50	89.10	0.18	7717	LIQ165TEEX
Teloschistes exilis	38.46	8275	5.97	87.40	0.41	8106	LIQ147TEX
Teloschistes flavicans	28.44	807	71.23	89.57	0.00	7615	LIQ166TFLAV
Teloschistes hypoglaucus	23.07	1856	18.19	88.51	0.12	7577	LIQ164TEHY
Variospora cf. flavescens	22.96	4566	9.42	82.59	0.06	6925	LIQ245VASP
Variospora fuerteventurae	32.90	2753	19.21	86.93	0.18	7013	LIQ242VAFU
Variospora sp. 1	33.87	351	157.49	91.21	0.00	7136	LIQ251GYDE
Variospora velana	34.61	2938	17.99	89.86	0.00	7604	LIQ163CAVE
Wetmoreana brouardii	31.50	2877	20.96	87.40	0.18	8294	LIQ189CABR
Wetmoreana decipioides	30.57	1139	40.14	89.51	0.00	7901	LIQ185CASP
Wetmoreana sp. 5	27.98	2054	29.00	89.16	0.00	7974	LIQ186CAVE
Xanthocarpia lactea	27.21	952	45.30	88.80	0.23	7745	LIQ231CALA
Xanthocarpia marmorata	43.48	7773	10.81	94.67	0.12	8013	LIQ223CAMA
Xanthocarpia ochracea	27.37	639	70.91	89.51	0.12	7527	LIQ230CAOCH
Xanthomendoza concinna	23.37	7038	5.31	78.78	0.41	7706	LIQ168XMCO
Xanthomendoza sp. 1	33.42	201	278.10	90.86	0.00	7644	LIQ169XMCO
Xanthomendoza sp. 2	50.51	9473	8.80	90.15	0.12	7611	LIQ200XMHS

Xanthomendoza weberi	23.46	1548	22.93	89.68	0.18	7784	LIQ201XMSP
Xanthomendoza weberi	23.93	1708	20.71	91.50	0.18	7909	LIQ202XMFA

Comparing trimmed and untrimmed alignments showed that both datasets produced the same topology (Supplementary Figure A.5).

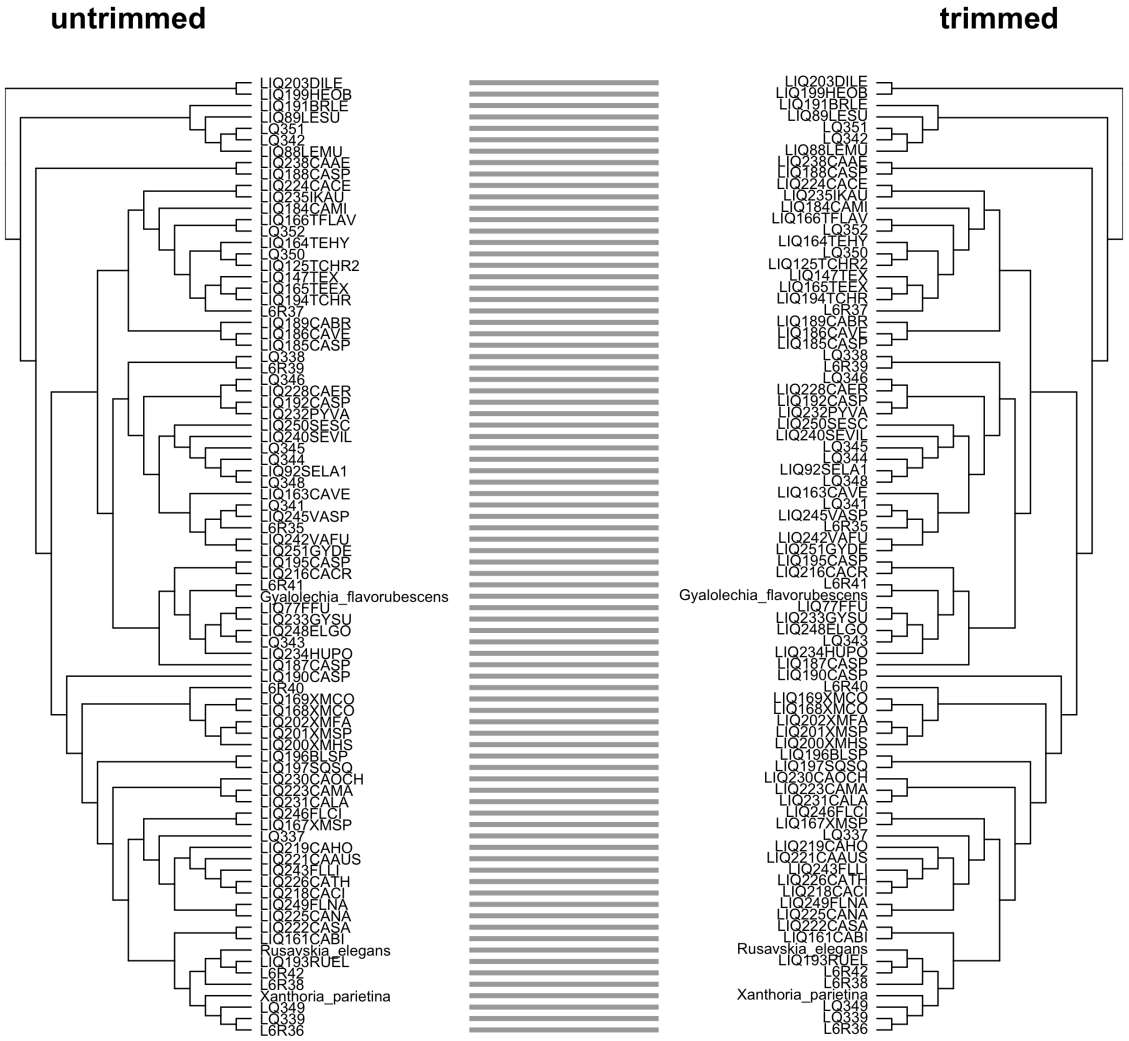


Figure A.5: Tanglegram comparing topology of ML phylogenies inferred from trimmed (right) and untrimmed (left) alignments.

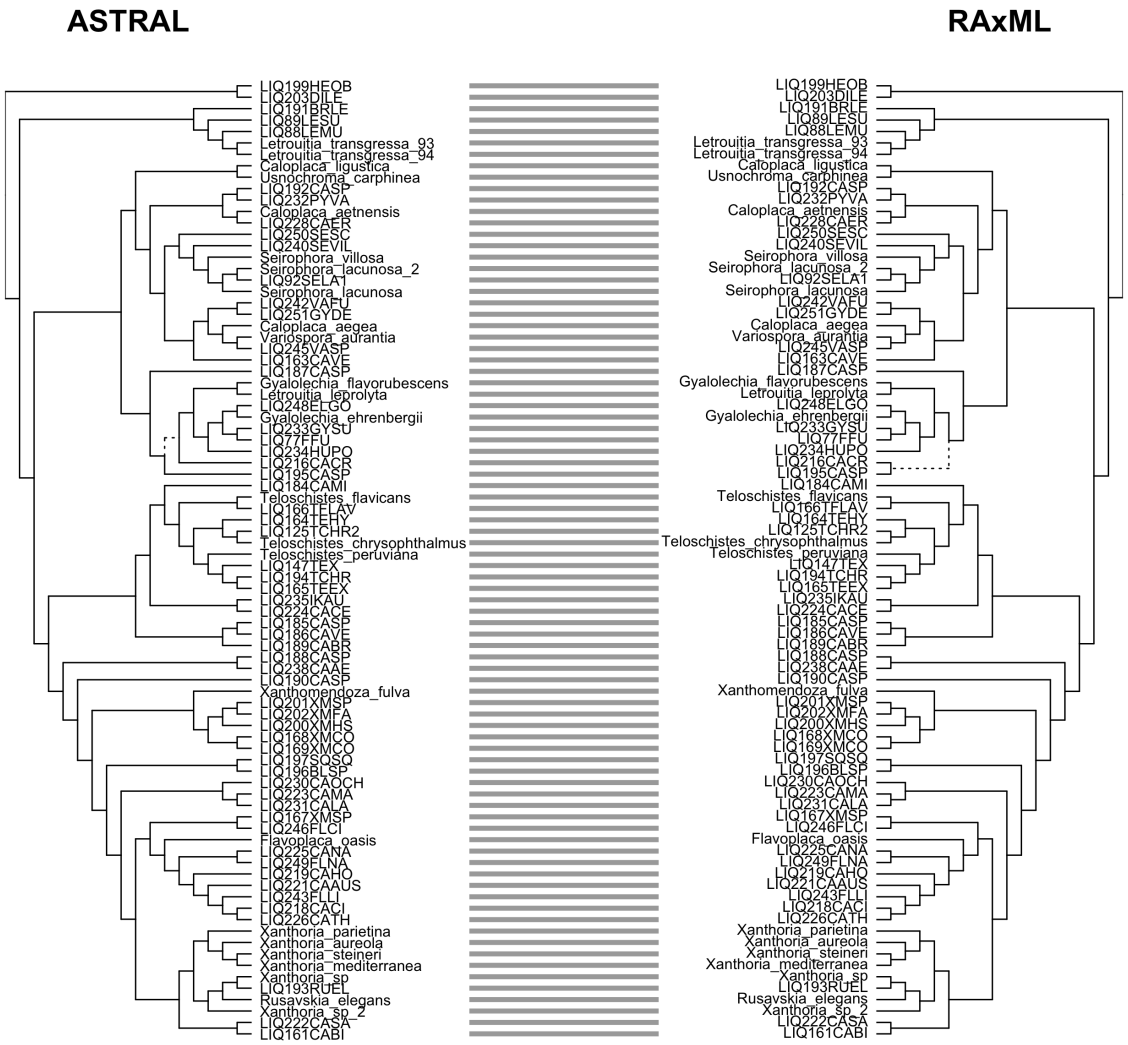


Figure A.6: Tanglegram comparing topology of the concatenated Maximum Likelihood (ML) phylogenomic tree inferred by RAxML (right) with the coalescent-based ASTRAL tree (left). Dashed branches show clades not found in the other phylogeny.

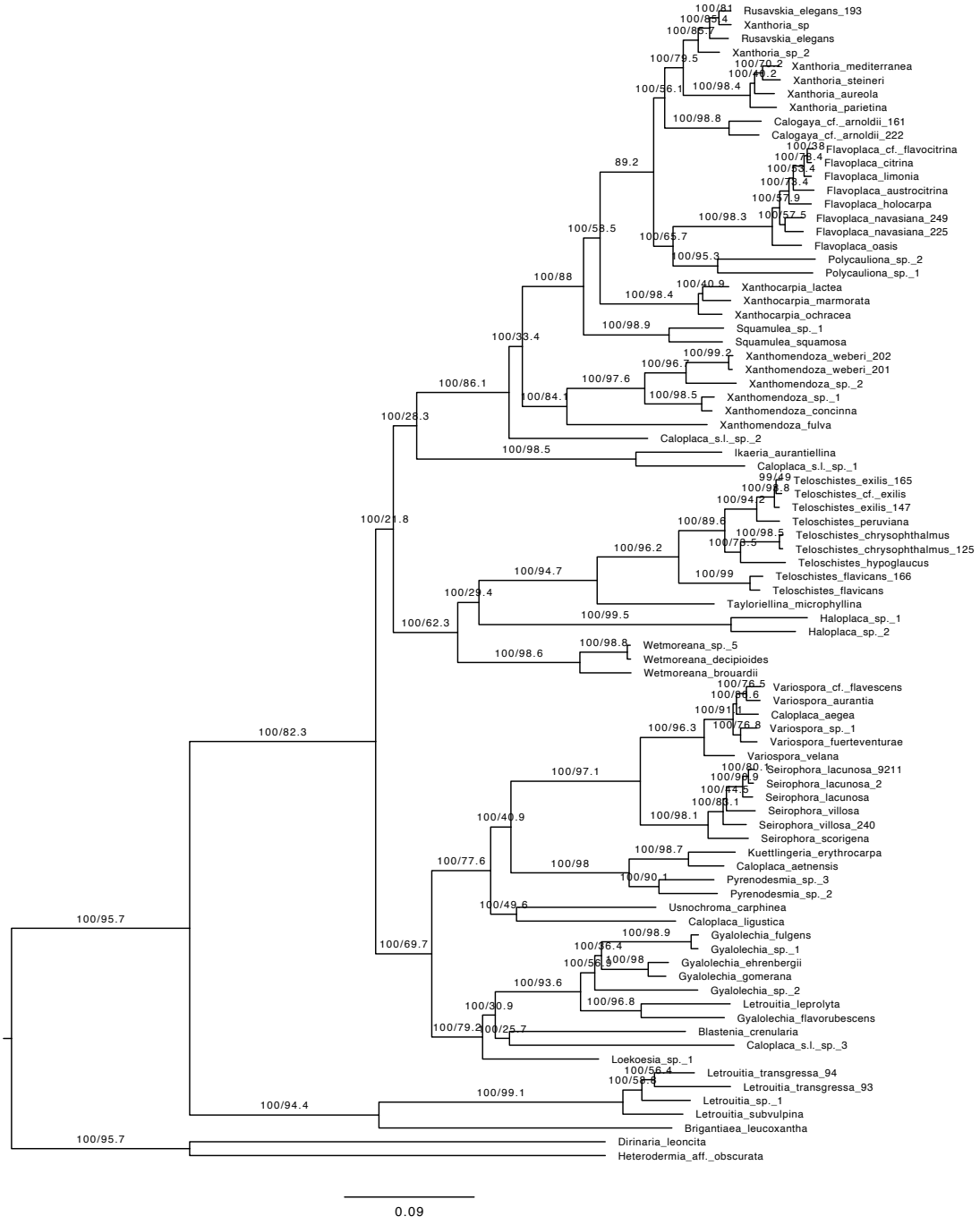


Figure A.7: 84-taxon genome-scale tree inferred by RAxML using 1254 loci. Support values are bootstraps and gene concordance factors (gCF)

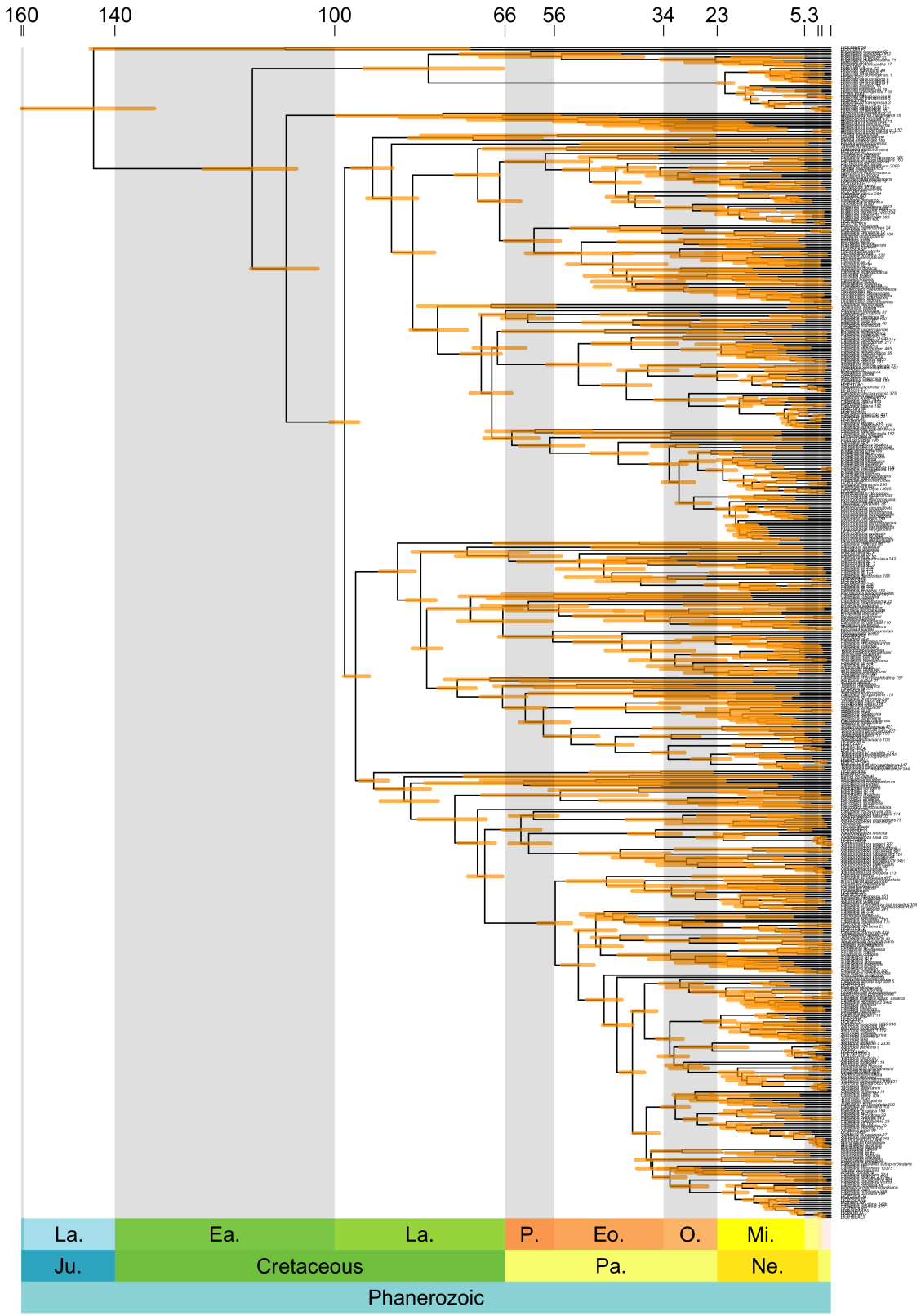


Figure A.8: Time-calibrated phylogeny of 550 *Teloschistales* taxa and *Caliciales* outgroup. Nodes are positioned on mean MCMCTree posterior estimates and orange bars show 95% highest posterior density. Ages shown in millions of years before present

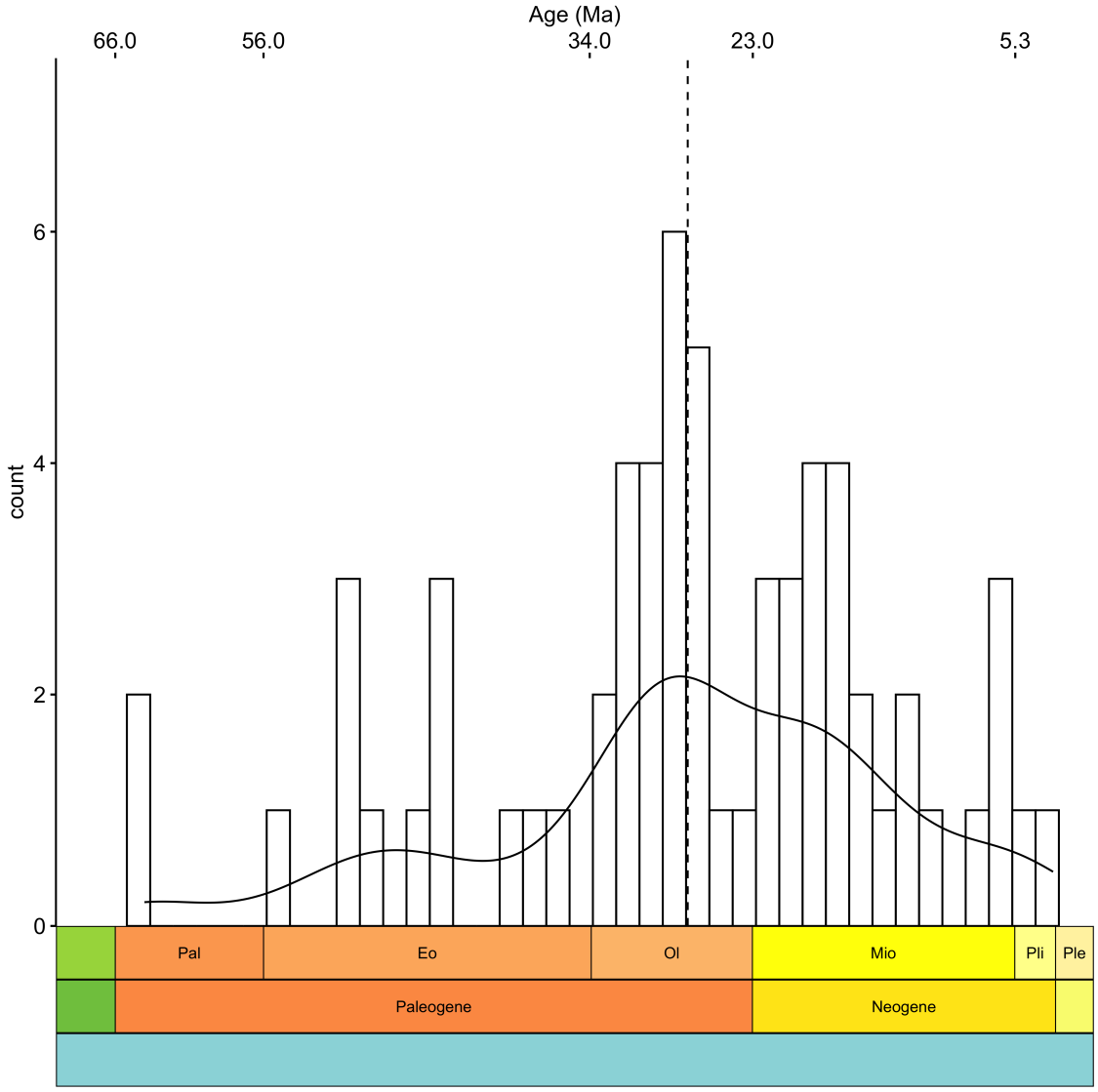


Figure A.9: Histogram showing the node ages of *Teloschistales* genera in Figure 3.2 as estimated by MCMCTree. Curved line is density plot, dashed line shows mean.

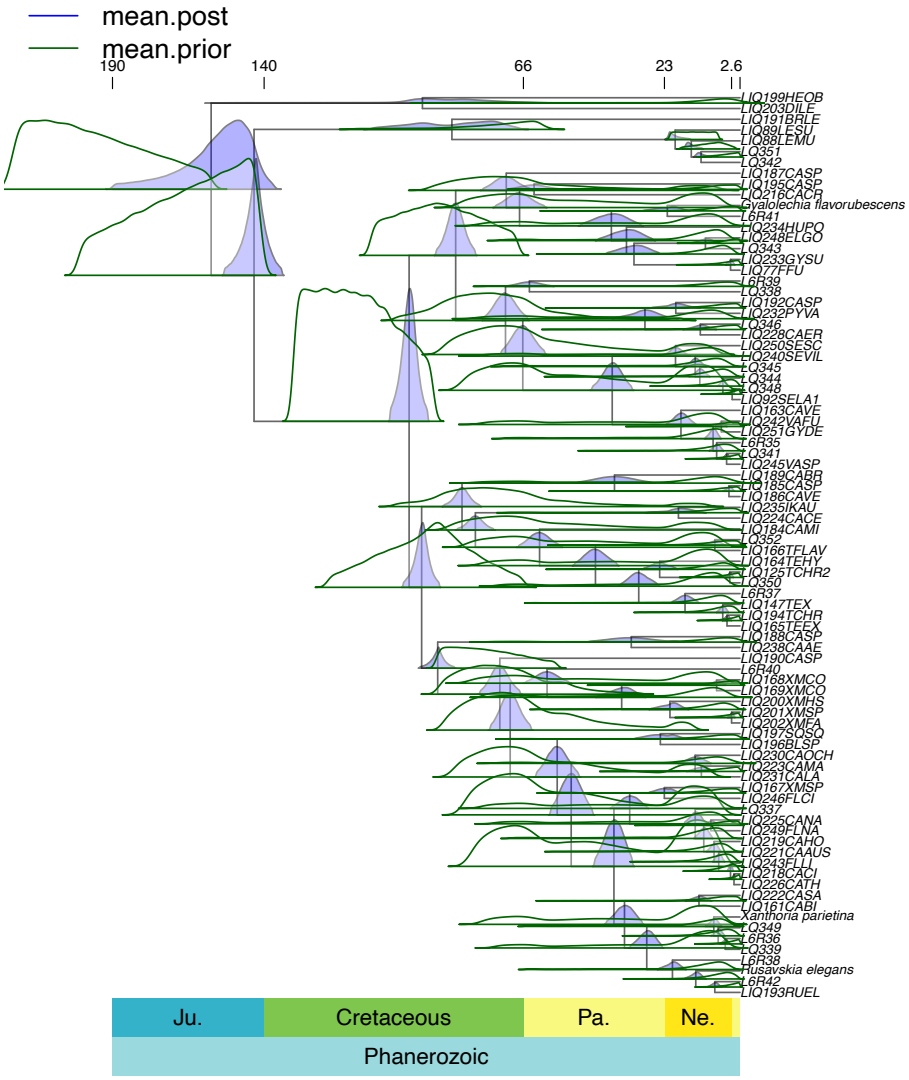


Figure A.10: Mean posterior vs mean prior node age estimates for the 84-taxon genome-scale phylogeny estimated with MCMCTree under a general brownian motion relaxed clock model.

A.3 Chapter 4 Supplementary Materials

Table A.7: Collection information for eight newly sequenced lichens in Chapter 4.

Taxon	Collection information (Locality, Coll., Date)
Coniocarpon cinnabarinum	Veracruz, Mexico, GG, 29/08/2020
Coniocarpon cinnabarinum	Veracruz, Mexico, GG, 19/08/2020
Protoblastenia rupestris	Coahuila, Mexico, AH, 20/07/2020
Pseudevernia consocians	Coahuila, Mexico, AH, 26/12/2020
Psora crenata	Northern Cape, South Africa, ET, 10/10/2014
Pyrenula sp. 1	Veracruz, Mexico, GG, 06/05/2020
Trypethelium eluteriae	Veracruz, Mexico, GG, 08/12/2019
Trypethellium subeluteriae	Mexico, missing, missing

Table A.8: Assembly statistics for eight additional non-*Teloschistales* mycobiont genomes

Taxon	Assembly size (Mbp)	Scaffolds	N50 (K)	BUSCO %	Dup. BUSCO %	Genes	Accession
Coniocarpon cinnabarinum	26.33	655	89.80	90.86	0.12	7913	LIQ171ARSP
Coniocarpon cinnabarinum	25.99	725	98.02	90.45	0.06	7721	LIQ211ARSP
Protoblastenia rupestris	25.36	3814	15.12	83.70	1.11	7919	LIQ213PRRU
Pseudevernia consocians	33.79	485	126.33	87.05	0.23	9778	LIQ198PSCO
Psora crenata	24.77	1663	28.52	90.15	0.18	6514	LIQ215PSCR
Pyrenula sp. 1	29.21	307	306.52	87.57	0.06	9330	LIQ210ARCI
Trypethelium eluteriae	29.04	1371	32.57	90.21	0	9187	LIQ180TREL
Trypethellium subeluteriae	29.28	251	206.44	92.67	0.12	8904	LIQ214TRSU

Table A.9: AIC results comparing the fit of three alternative evolutionary models for BGC and metabolite number per sample. BM = Brownian Motion, EB = Early Burst Model, OU = Orstein-Uhlenberg.

Dataset	Trait	BM	EB	OU
All Taxa	bgcs	597.37	599.37	579.95
	molecules	857.62	859.62	822.26
	+ve ions	1183.34	1185.34	1135.64
	-ve ions	914.43	916.43	889.23
	+ve GNPS	1148.93	1150.93	1099.87
	-ve GNPS	863.74	865.74	833.1
Teloschistales	bgcs	509.33	511.33	501.29
	molecules	729.32	731.32	715.53
	+ve ions	1011.27	1013.27	990.3
	-ve ions	768.84	770.84	763.02
	+ve GNPS	981.2	983.2	959.62
	-ve GNPS	723.86	725.86	717.06

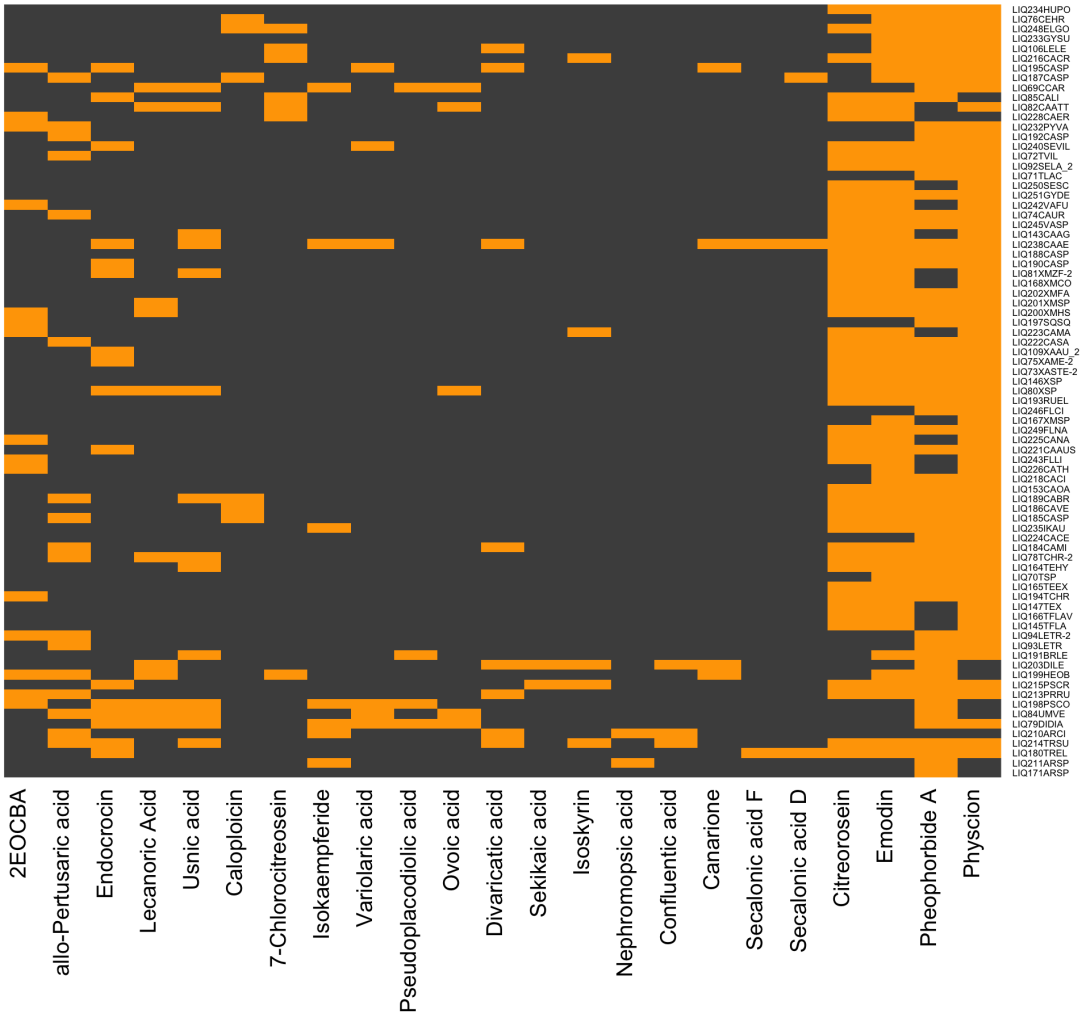


Figure A.11: Presence-absence matrix of annotated secondary metabolites from GNPS feature-based molecular networking. Heatmap columns are ordered according to hierarchical clustering of metabolite presence-absence patterns. Rows represent lichen samples ordered by maximum likelihood phylogeny. Presence = orange, absence = grey.

Table A.10: Results of *hOUwie* analysis jointly modelling the number of anthraquinone-like GNPS metabolite features per sample (continuous trait) and the evolution of individual genes within the anthraquinone BGC (discrete traits). Continuous trait models are BM1 = Brownian Motion with one evolutionary rate across discrete character states, BMV = BM with variable evolutionary rate across discrete character states, OU1 = Orstein-Uhlenbeck with one optimal trait value across discrete character states, OUM = OU with multiple optimal trait values across discrete character states. '_cid' shows results for models allowing variable parameter models without variation being linked to an observed character. Columns are number of model parameters, log-likelihood, discrete trait likelihood, continuous trait likelihood and weighted BIC. Best-fitting models are highlighted by an asterisk after weighted BIC.

gene	Model	np	lnLik	DiscLik	ContLik	BICwt
EthD	BM1	8	-10000000057.6724	-57.6693	-9999999995.3948	0.8792*
	OU1	9	-10000000057.6568	-57.7300	-9999999995.3948	0.1067

	BMV	10	-10000000057.6745	-57.7110	-9999999995.3948	0.0125
	OUM	11	-10000000057.6786	-57.7253	-9999999995.3948	0.0015
	BMV_cid	17	-10000000058.3058	-58.5933	-9999999995.3948	0.0000
	OUM_cid	18	-10000000057.9059	-58.8247	-9999999995.3948	0.0000
PKSI	BM1	4	-10000000009.9792	-9.9935	-9999999995.3948	0.7974*
	OU1	5	-10000000009.9771	-9.9873	-9999999995.3948	0.0955
	BMV	5	-10000000009.9752	-9.9804	-9999999995.3948	0.0957
	OUM	6	-10000000009.9782	-9.9819	-9999999995.3948	0.0114
	BMV_cid	9	-10000000010.1126	-10.1337	-9999999995.3948	0.0000
	OUM_cid	10	-10000000010.1060	-10.1359	-9999999995.3948	0.0000
MBL	BM1	4	-10000000009.9764	-9.9903	-9999999995.3948	0.8117*
	OU1	5	-10000000010.0581	-10.3640	-9999999995.3948	0.0894
	BMV	5	-10000000009.9818	-9.9910	-9999999995.3948	0.0965
	OUM	6	-10000000011.5554	-11.5629	-9999999995.3948	0.0024
	BMV_cid	9	-10000000010.1038	-10.1037	-9999999995.3948	0.0000
	OUM_cid	10	-10000000010.0952	-10.1222	-9999999995.3948	0.0000
ABC	BM1	8	-10000000027.2289	-27.3051	-9999999995.3948	0.8886*
	OU1	9	-10000000027.3120	-27.3246	-9999999995.3948	0.0977
	BMV	10	-10000000027.2640	-27.2973	-9999999995.3948	0.0123
	OUM	11	-10000000027.3084	-27.2892	-9999999995.3948	0.0014
	BMV_cid	17	-10000000028.1925	-28.2328	-9999999995.3948	0.0000
	OUM_cid	18	-10000000029.5994	-29.6664	-9999999995.3948	0.0000
MethylT	BM1	8	-10000000053.2690	-53.3935	-9999999995.3948	0.9300*
	OU1	9	-10000000054.0567	-54.1004	-9999999995.3948	0.0506
	BMV	10	-10000000052.9965	-53.0923	-9999999995.3948	0.0174
	OUM	11	-10000000053.0332	-53.0878	-9999999995.3948	0.0020
	BMV_cid	17	-10000000056.0476	-56.1152	-9999999995.3948	0.0000
	OUM_cid	18	-10000000053.8509	-65.3103	-9999999995.3948	0.0000
GammaG	BM1	4	-10000000028.3807	-28.4936	-9999999995.3948	0.8098*
	OU1	5	-10000000029.2591	-29.3227	-9999999995.3948	0.0402
	BMV	5	-10000000028.0043	-28.4102	-9999999995.3948	0.1410
	OUM	6	-10000000028.6398	-28.8453	-9999999995.3948	0.0089
	BMV_cid	9	-10000000029.7238	-85.8764	-9999999995.3948	0.0000
	OUM_cid	10	-10000000032.1332	-35.9231	-9999999995.3948	0.0000
AAperm	BM1	8	-10000000034.7179	-35.9315	-9999999995.3948	0.8441*
	OU1	9	-10000000034.4945	-34.4942	-9999999995.3948	0.1261
	BMV	10	-10000000034.2203	-38.0311	-9999999995.3948	0.0198
	OUM	11	-10000000032.7865	-33.1424	-9999999995.3948	0.0099
	BMV_cid	17	-10000000044.8381	-47.7179	-9999999995.3948	0.0000
	OUM_cid	18	-10000000035.0516	-35.3173	-9999999995.3948	0.0000
Oreduct	BM1	8	-10000000042.8859	-43.0857	-9999999995.3948	0.8875*
	OU1	9	-10000000042.9552	-43.0979	-9999999995.3948	0.0990
	BMV	10	-10000000042.9104	-43.6546	-9999999995.3948	0.0124
	OUM	11	-10000000043.1543	-43.5590	-9999999995.3948	0.0012
	BMV_cid	17	-10000000047.8353	-47.8109	-9999999995.3948	0.0000
	OUM_cid	18	-10000000042.1756	-43.4813	-9999999995.3948	0.0000
KinaseP	BM1	8	-10000000059.7774	-61.7400	-9999999995.3948	0.8733*
	OU1	9	-10000000059.6597	-61.7784	-9999999995.3948	0.1174
	BMV	10	-10000000060.2270	-60.5184	-9999999995.3948	0.0080
	OUM	11	-10000000059.9061	-60.0677	-9999999995.3948	0.0013
	BMV_cid	17	-10000000062.2410	-64.9609	-9999999995.3948	0.0000
	OUM_cid	18	-10000000061.9763	-62.1531	-9999999995.3948	0.0000
DUF	BM1	8	-10000000046.4485	-46.8122	-9999999995.3948	0.6465*
	OU1	9	-10000000045.0154	-45.1721	-9999999995.3948	0.3239

BMV	10	-10000000045.2932	-45.5054	-9999999995.3948	0.0293
OUM	11	-10000000047.6540	-47.9858	-9999999995.3948	0.0003
BMV_cid	17	-10000000051.3846	-55.4848	-9999999995.3948	0.0000
OUM_cid	18	-10000000046.8930	-47.7650	-9999999995.3948	0.0000

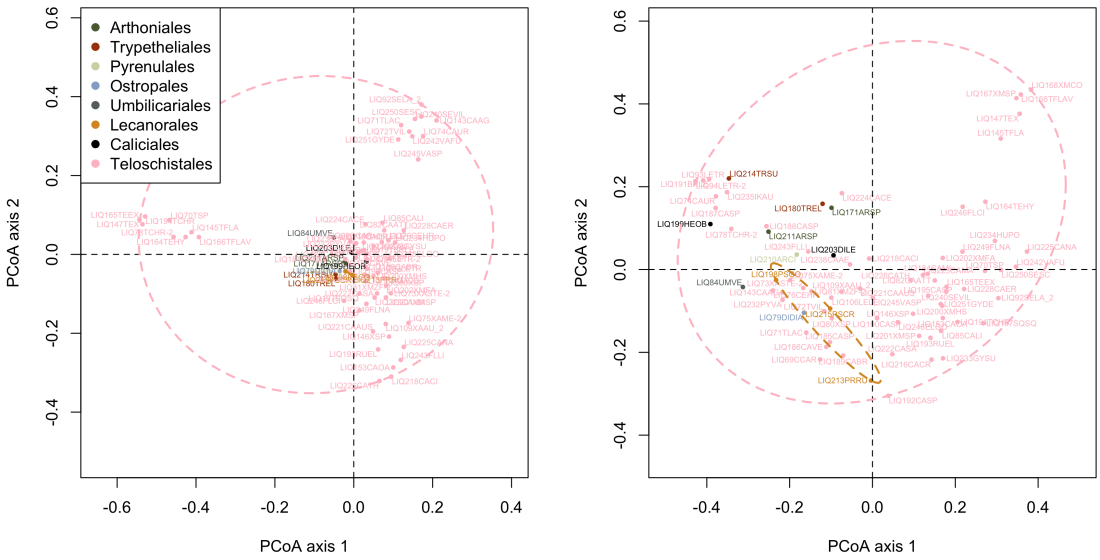


Figure A.12: PCoA plots of BGC family (left) and computed neutral metabolites (right) presence-absence patterns. Colours represent taxonomic orders with ellipses drawn around groups with more than three points

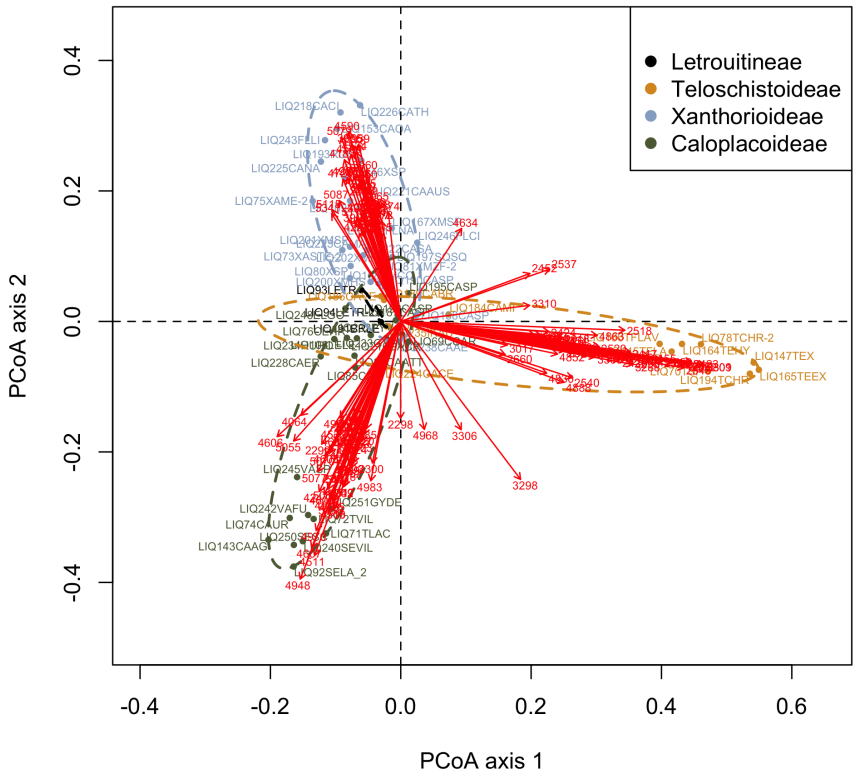


Figure A.13: PCoA plot of BGC family presence-absence patterns seen in Figure 4.2 with individual BGCs fitted as environmental variables (red arrows). Only environmental variables with a correlation p-value<0.01 are shown. Colours represent taxonomic subfamilies of *Teloschistaceae* and outgroup suborder *Letrouitineae*, with ellipses drawn around groups.

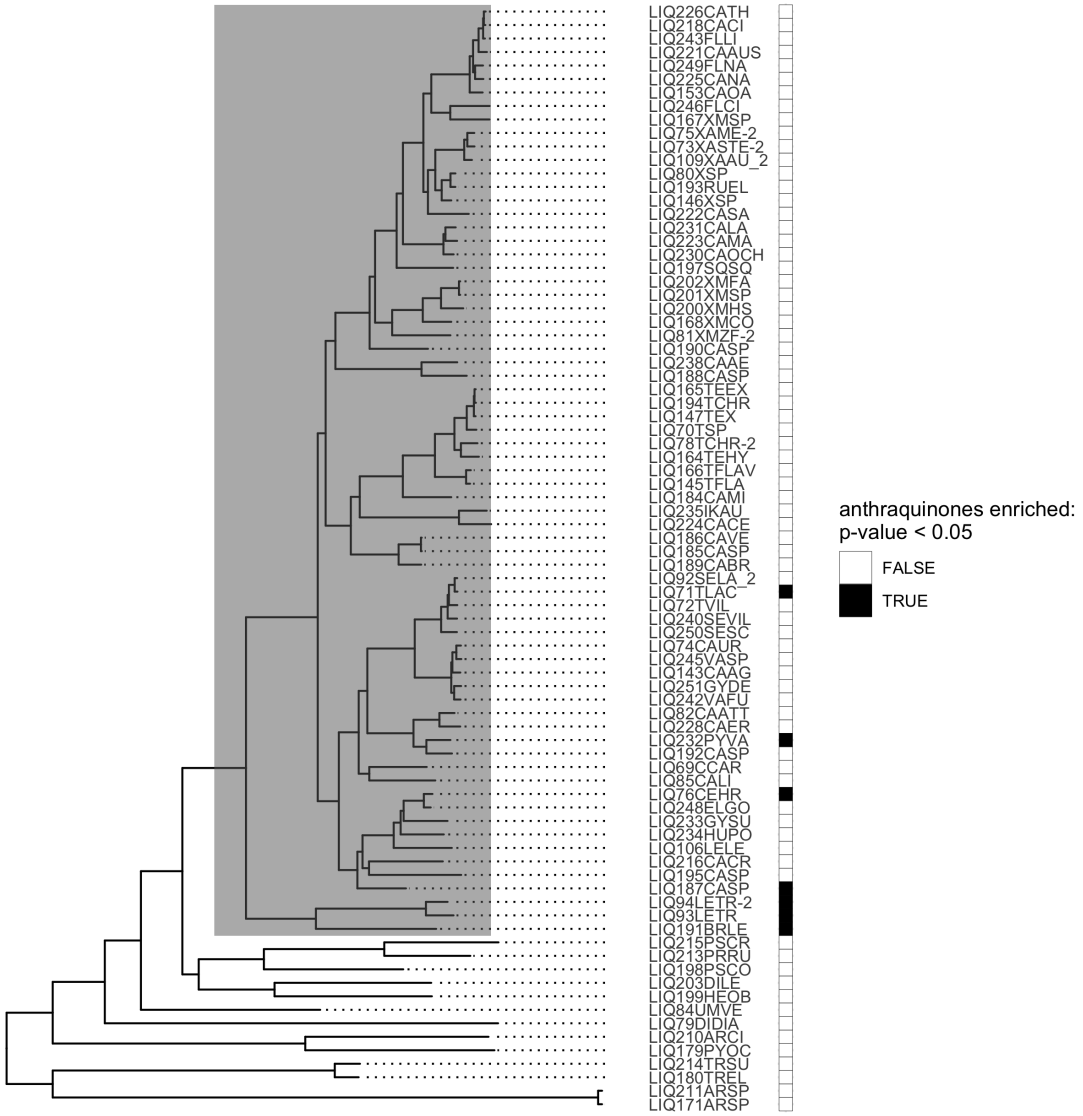


Figure A.14: Results of enrichment analysis testing whether *Teloschistales* taxa are more enriched in anthraquinone GNPS subnetworks. Black = taxon significantly enriched, white = not significantly enriched. Phylogeny is IQTree maximum likelihood phylogeny of 1351 single-copy protein-coding genes. *Teloschistales* highlighted with grey square

A.4 Chapter 5 Supplementary Materials

Table A.11: Lichen samples used for LCMS emodin and parietin detection/ quantification.

Sample number	Species	Collection information (Locality, Coll., Date
LIQ110CASP	*Caloplaca* sp. 1	Isola del Giglio, Italy, TBL, 13.09.2019
LIQ79DIDIA	*Diploschistes diacapsis*	Catalunya, Spain, XL, AA, 04.09.2018
LIQ125TCHR	*Teloschistes chrysophthalmus*	Catalunya, Spain, XL, AA, 02.04.2019
LIQ254RAFAR	*Ramalina farinacea*	Surrey, UK, TBL, 02.11.2021
LIQ255PHSP	*Physcia* sp. 1	Surrey, UK, TBL, 02.11.2021
LIQ256PHYSP	*Physcia* sp. 2	Surrey, UK, TBL, 02.11.2021
LIQ258PHYSP	*Physcia* sp. 3	Surrey, UK, TBL, 02.11.2021
LIQ260LEEL	*Lecidella elaeochroma*	Surrey, UK, TBL, 02.11.2021
LIQ261PHSP	*Physcia* sp. 4	Surrey, UK, TBL, 02.11.2021
LIQ262LEEL	*Lecidella elaeochroma*	Surrey, UK, TBL, 02.11.2021
LIQ277XAPA-P3	*Xanthoria parietina*	Poland, KW
LIQ2785	*Flavoplaca citrina*	Poland, KW
LIQ284XAPA-P4	*Xanthoria parietina*	Poland, KW
LIQ287RUEL7	*Rusavskia elegans*	Poland, KW
LIQ286CASU-3	*Haloplaca suaedae*	Dorset, UK, TL, EG, BE, 13.03.2022
LIQ287CACE	*Caloplaca cerina*	Dorset, UK, TL, EG, BE, 13.03.2022
LIQ304CAPO-1	*Caloplaca pollinii*	Dorset, UK, TL, EG, BE, 13.03.2022
LIQ311CACI	*Flavoplaca citrina* gr.	Dorset, UK, TL, EG, BE, 13.03.2022

Note: For *Teloschistes chrysophthalmus* a different strain had to be used for the symbiotic lichen thallus sample as the original LIQ125TCHR strain was unfortunately lost. We therefore used strain LIQ78TCHR which was collected at the same time and sample location as LIQ125TCHR

Table A.12: Collection information for four newly sequenced lichens in Chapter 5.

Taxon	Collection information (Locality, Coll., Date)
Circinaria calcarea	Switzerland, EG, PN, NH, AG, UA, US, 14.08.2006
Pyrenula ochraceoflavens	Veracruz, Mexico, GG, 08.12.2020
Cornicularia normoerica	Catalunya, Spain, XL, 21.09.2018
Protoblastenia testacea	Catalunya, Spain, XL, AA, 04.09.2018

Table A.13: Genome assembly statistics for four additional non-*Teloschistales* mycobiont genomes

Taxon	Assembly size (Mbp)	Scaffolds	N50 (K)	BUSCO %	Dup. BUSCO %	Genes	Accession
Circinaria calcarea	31.50	5513	9.07	87.05	1.76	9689	LIQ162CACH
Pyrenula ochraceoflavens	26.70	189	313.36	90.62	0.23	9541	LIQ179PYOC
Cornicularia normoerica	40.55	7206	12.47	86.58	0.06	8268	LIQ83CONO
Protoblastenia testacea	24.30	1839	31.56	89.62	0.23	7078	LIQ108PROTE

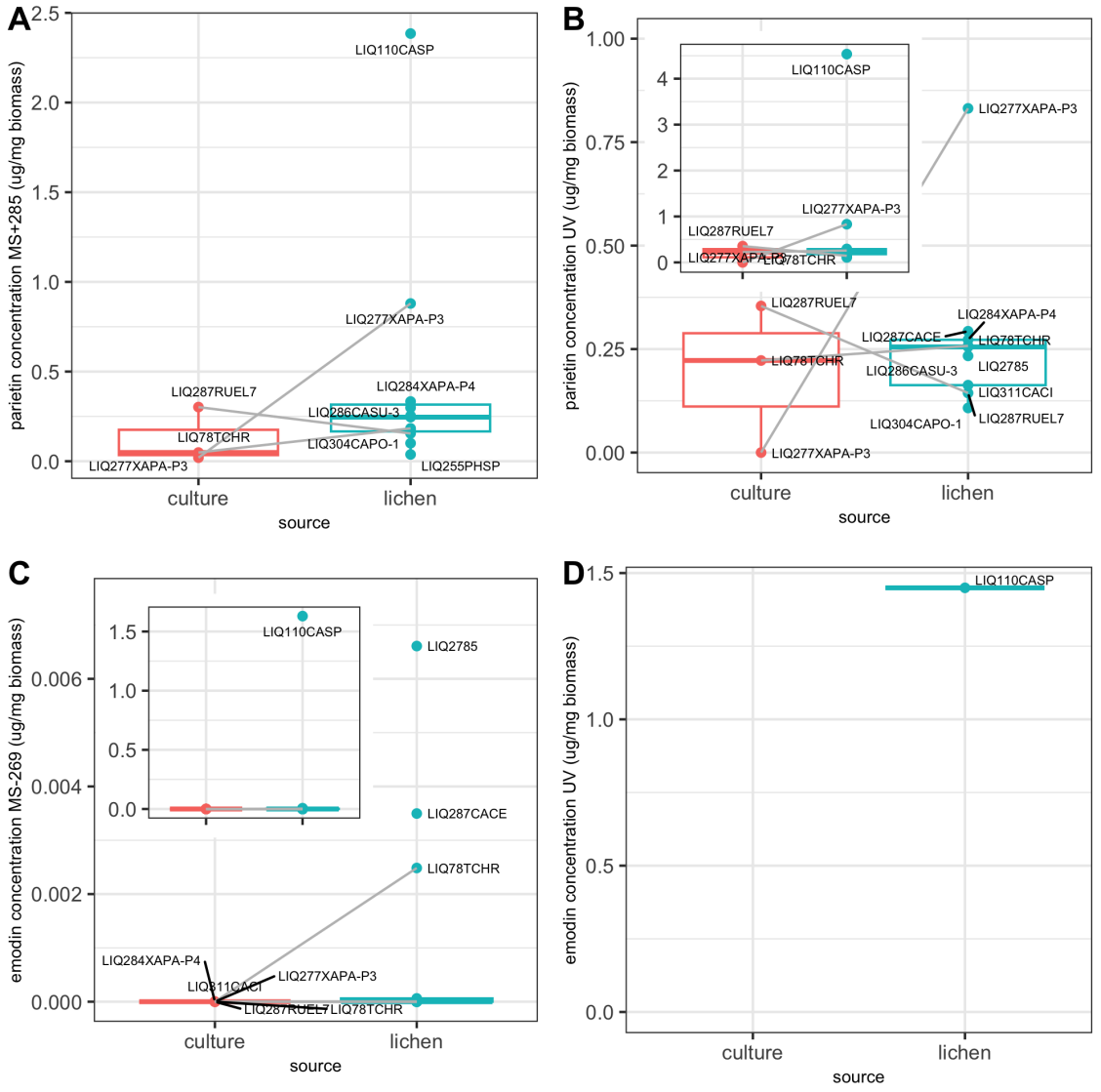


Figure A.15: Concentrations of emodin and parietin per mg of cultured aposymbiotic lichen mycobiont and whole lichen thallus. Concentrations were calculated from standard curves of pure compound serial dilutions. A) Concentration of parietin in cultures (orange) and lichen thalli (blue) calculated from peak areas of MS+ peak at 285 m/z. B) Same as A but calculated using peak areas of UVmax at 440nm. Main plot shows just concentrations between 0 and 1.0 ug/mg fungal biomass. Inset shows also outlier LIQ110CASP. C) Concentrations of emodin calculated from peak areas of MS- speak at 269 m/z. Main plot shows just concentrations between 0 and 0.0075 ug/mg fungal biomass. Inset shows also outlier LIQ110CASP. D) Same as C but calculated using peak areas of UVmax at 435nm. Concentrations below limit of quantification are shown as 0 values

A.5 *Ramalina farinacea* genome note (doi:10.1093/gbe/evad074)

GBE

First Whole-Genome Sequence and Flow Cytometry Genome Size Data for the Lichen-Forming Fungus *Ramalina farinacea* (Ascomycota)

Theo Llewellyn [1,2,3,*], Sahr Mian[1], Rowena Hill[1,4], Ilia J. Leitch [1], and Ester Gaya[1]

[1]Comparative Fungal Biology, Royal Botanic Gardens, Kew, Richmond, United Kingdom
[2]Department of Life Sciences, Imperial College London, Ascot, Berkshire, United Kingdom
[3]Science and Solutions for a Changing Planet Doctoral Training Partnership, Grantham Institute, Imperial College London, South Kensington, London, United Kingdom
[4]School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom

*Corresponding author: E-mails: t.llewellyn@kew.org; t.llewellyn19@imperial.ac.uk.

Accepted: 27 April 2023

Abstract

Lichen-forming fungi are a diverse and ecologically important group of obligate mutualistic symbionts. Due to difficulties with maintaining them in culture and their extremely slow growth, lichenologists are increasingly opting for metagenomic sequencing followed by symbiont genome separation using bioinformatic pipelines. However, without knowing the true genome size of the lichen-forming fungus, we cannot quantify the completeness of the genome assembly and the efficacy of the bioinformatic filtering. To address this issue, we report here the first whole-genome assembly for the lichen-forming fungus *Ramalina farinacea* (L.) Ach. sequenced with Oxford Nanopore long-read technology alongside direct measurements of its genome size using flow cytometry. The assembly showed high contiguity (N50 = 1.55 Mb) and gene set completeness (BUSCO = 95.8%). The highly robust genome size of 33.61 Mb/1C (coefficients of variation = 2.98) that was obtained showed our assembly covered 97% of the entire genome. Our results demonstrate that accurate genome size measurements can be obtained directly from lichen thalli and used to provide a benchmark for assessing true cytometric completeness of metagenome-derived assemblies.

Key words: genome size, lichen, long-read sequencing, metagenomics, Ramalinaceae, symbiosis.

Significance

Whole-genome sequencing of obligate fungal symbionts is increasingly possible due to advances in sequencing technology and metagenomics methods. Metagenomic pipelines have taxon filtering steps to remove nontarget organism sequences. However, these steps can be inaccurate as they rely on indirect genome assembly metrics to determine genome completeness and contamination, which are often underestimates. Here, we sequence a lichen-forming fungal symbiont genome using long-read sequencing and concurrently measure for the first time its genome size directly using flow cytometry to ensure assembly completeness and validate metagenomic filtering. This study provides a roadmap for incorporating cytometric analysis into genome-sequencing projects of lichen-forming fungi, an approach we believe is necessary for producing high-quality, uncontaminated genomes from metagenomic data.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

Introduction

Fungal symbionts make up a significant proportion of the kingdom *Fungi*, with symbiotic lifestyles such as lichens, mycorrhizae, and entomopathogens spanning all major lineages (Peay et al. 2016). Given their prevalence, sequencing the genomes of these fungi is a key step in understanding the evolution of symbiotic lifestyles (Miyauchi et al. 2020). One such group is lichen-forming fungi, of which approximately 19,000 species have been described (Lücking et al. 2017). Most lichen-forming fungi (mycobionts) cannot be cultured axenically, which makes standard laboratory practices for whole-genome sequencing challenging. As a result, metagenomic sequencing is increasingly used as an alternative in lichen genomics studies (e.g., McKenzie et al. 2020; Tzovaras et al. 2020; Allen et al. 2021; Tagirdzhanova et al. 2021). When assembling fungal symbiont genomes from metagenomic data, metrics such as assembly size and BUSCO scores are used as benchmarks for assessing assembly completeness and to ensure that sequences from the other symbionts have been removed. However, bioinformatic genome size estimates from assembly size or k-mer-based approaches can differ significantly from direct cytometric genome size measurements. This disparity can be due to the fact that BUSCOs measure only gene set completeness and cannot quantify the completeness of noncoding (repeat) regions (Le Cam et al. 2019; Hill et al. 2021). As a result, seemingly complete assemblies may be missing significant portions of the nongenic regions of the genome.

So far, no lichen-genome-sequencing studies have incorporated cytogenetic estimates of genome size made using flow cytometry. Consequently, we do not know yet how well current sequencing and assembly pipelines are performing for lichen metagenomic data. To address this issue, we concurrently performed, for the first time, long-read genome sequencing and flow cytometry genome size analysis for the lichen-forming fungus *Ramalina farinacea* (fig. 1*a*), a widespread fruticose (bushy) species in the family *Ramalinaceae* (*Lecanorales*, *Ascomycota*). We chose this species given its abundance, large thalli, and lack of a whole-genome sequence. *Ramalina farinacea* is a globally distributed epiphytic, lichen-forming fungus. It associates with a high diversity of algal symbionts predominantly from the genus *Trebouxia* (Molins et al. 2021). It is also chemically diverse, being able to produce various depside and depsidone secondary metabolites (Stocker-Wörgötter et al. 2004). Only four other *Ramalinaceae* genomes have been sequenced to date; two *Ramalina* species using short-read Illumina sequencing, and two *Bacidia* species using Oxford Nanopore long-read sequencing (Allen et al. 2021; Gerasimova et al. 2022).

Results and Discussion

Here, we present the genome sequence for the lichenized fungus *R. farinacea* (fig. 1*b*). This is the first published whole-genome sequence for this species and the first long-read assembly for the genus. The taxonomic identification of the sample was confirmed with a blast search of the genome-derived full ITS sequence against the UNITE database (Nilsson et al. 2019) and T-BAS phylogenetic placement (Carbone et al. 2017). Both approaches confirmed the sample to be the target species *R. farinacea* (see Supplementary Material online). Our chosen metagenomic pipeline (Llewellyn et al. 2023) successfully isolated the mycobiont reads from the metagenomic data set (fig. 1*c* and *d*). K-mer-based profiling of unprocessed mycobiont reads estimated the genome size at 31.53 Mb. The mycobiont sequence data were then assembled into 44 contigs with an assembly size of 32.74 Mb, of which 29 contigs constituted 99% (32.56 Mb) of the assembly. This falls within the range of published assembly sizes for species belonging to *Lecanoromycetes*. The assembly size was approximately 5 Mb larger than the two Illumina published *Ramalina* genomes and comparable with the *Bacidia* long-read genomes (table 1). The assembly was highly contiguous (N50 = 1.55 Mb) and the longest scaffold measured 2.45 Mb. This is at least an order of magnitude higher than the contiguity scores for the other two published *Ramalina* assemblies, confirming the trend that long-read approaches significantly improve mycobiont assembly quality. N50 is comparable with the available long-read assemblies for the *Lecanoromycetes* genera *Bacidia*, *Physcia*, *Letharia*, and *Umbilicaria* (McKenzie et al. 2020; Tzovaras et al. 2020; Wilken et al. 2020; Allen et al. 2021; Singh et al. 2021; Gerasimova et al. 2022). The assembly had a BUSCO completeness score of 95.8% (table 1).

Of the 41 contigs over 5,000 base pairs long, 13 had telomere sequences at both ends possibly implying a contig length covering a full chromosome (fig. 1*e*); a further 20 contigs had telomere sequences at one end. This result suggests the *R. farinacea* genome produced here has at least 13 chromosomes, although karyotyping would be needed to confirm this.

Genome annotation predicted 8,765 proteins, and secondary metabolite biosynthetic gene cluster (BGC) analyses identified 52 BGCs; 25 polyketide synthases (PKSs), 16 nonribosomal peptide synthetases (NRPSs), 4 terpenes, 5 NRPS-PKS hybrids, and 1 indole cluster. *Ramalina farinacea* has a similar number of BGCs to *Ramalina intermedia* and *Ramalina peruviana* which have 54 and 43 clusters, respectively (table 1). The number of BGCs is significantly higher than in the two published *Bacidia* genomes, supporting a previous observation by Gerasimova et al. (2022). The aforementioned authors attribute this to the more diverse secondary chemistry profiles of *Ramalina* species which frequently produce multiple depside and depsidones, while only a single secondary metabolite has been observed per species in the two *Bacidia* species sequenced thus far. Our new data contributes to a growing *Ramalinaceae* data set that may help inform evolutionary trends in the

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

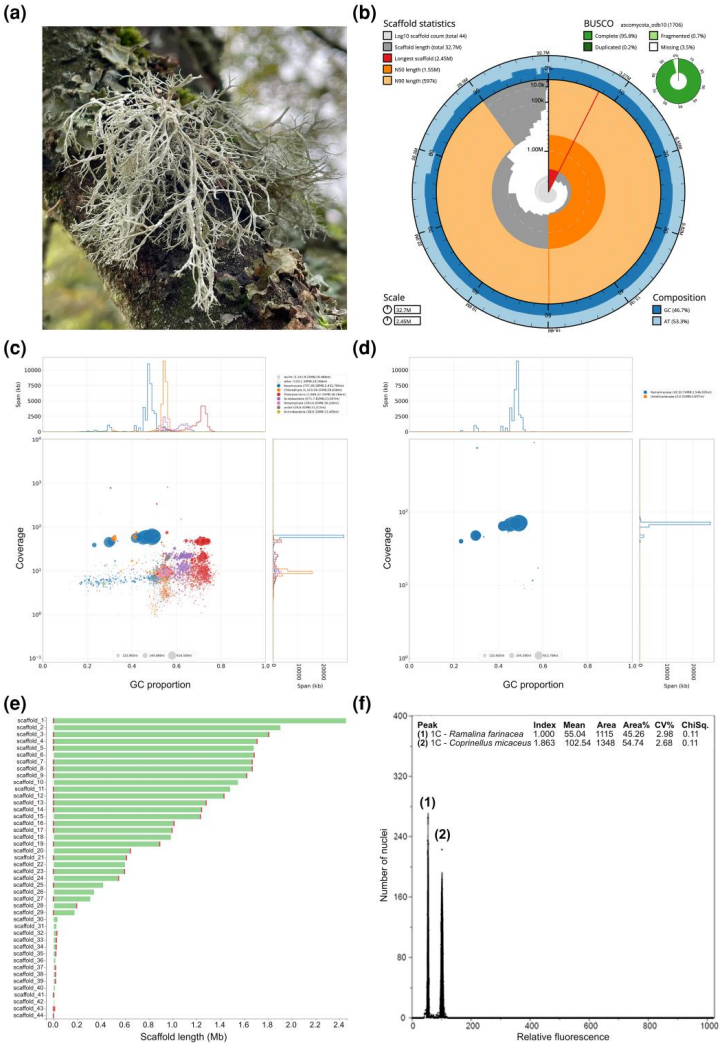


Fig. 1.—Genome assembly of *Ramalina farinacea*. (*a*) Thallus of *R. farinacea* growing on the *Quercus brantii* tree. (*b*) BlobToolKit Snailplot of mycobiont assembly showing contig length, N50/90, and BUSCO scores. (*c*) Metagenome blob plot before mycobiont filtering displaying GC-content against sequencing coverage for every contig in the metagenome assembly, with corresponding axis histograms. Color represents taxonomic identification via DIAMOND and BLAST+ searches. Circle diameter = contig length. Legend shows contig number, span, and N50 for each taxon. (*d*) Blob plot showing GC content versus coverage for isolated mycobiont genome after removing nonmycobiont contigs and reassembling (*e*) Plot showing the number and length of all 44 assembled scaffolds for the mycobiont assembly. Vertical red bars show the presence of telomere sequences (*f*) Example of a flow cytometry histogram used to estimate the genome size of *R. farinacea*. Peaks 1 and 2 correspond to propidium iodide–stained nuclei of *R. farinacea* and *Coprinellus micaceus* (internal calibration standard), respectively.

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

Table 1
Genome Assembly Statistics for *Ramalina Farinacea* (new to This Study) Compared With Published Ramalinaceae Genome Assemblies

	Assembly Size (Mb)	N50 (Mb)	No. Scaffolds/ Contigs (Mb)	BUSCO Completeness (%)	Predicted Proteins	BGCs	Bioproject Accession
Ramalina farinacea	32.74	1.55	44	95.8 (95.6)	8,765	52	PRJNA890456
Ramalina intermedia	26.25	0.273	198	95.8 (95.8)	7,355	54	PRJNA416400
Ramalina peruviana	26.99	0.04	1657	89.6 (89)	6,756	43	PRJNA358650
Bacidia gigantensis	33.12	1.81	24	93.3 (93.2)	8,400	35	PRJNA748063
Bacidia rubella	33.52	1.77	51	82.7 (82.6)	8,469	31	PRJNA821848

Note.—The BUSCO completeness percentages for assembled genomes from other studies were recalculated in this study to ensure comparability. Scores in brackets are single-copy BUSCOs. The BUSCO score for *Bacidia rubella* could be calculated only on the published protein set as the assembled genome has not been made publicly available. BGCs, secondary metabolite biosynthetic gene clusters.

family. For example, to understand why *Ramalina* genomes have more diverse BGC repertoires than *Bacidia*, and whether the ancestral state of the family is a diverse BGC repertoire which was then reduced in *Bacidia* or alternatively, an ancestral BGC-poor repertoire which then expanded in *Ramalina*.

The flow cytometry histogram produced by analyzing the combined sample of propidium iodide–stained nuclei from *R. farinacea* and the internal calibration standard *Coprinellus micaceus* showed two distinct peaks (fig. 1*f*). Peak 1 (mean relative fluorescence of 55.04) corresponds to nuclei from *R. farinacea*, whereas Peak 2 (mean relative fluorescence of 102.54) comprises nuclei from *C. micaceus*. Based on the following formula: (Mean relative fluorescence of species of interest [i.e., *R. farinacea*]/Mean relative fluorescence of calibration standard [i.e., *C. micaceus*]) × 1C-value of calibration standard (i.e., 62.62 Mb), the genome size of *R. farinacea* is estimated to be 33.61 Mb. Given the low coefficients of variation of both peaks in the flow histogram (i.e., 2.98% for the *R. farinacea* [peak 1], and 2.68% for *C. micaceus* [peak 2]), the genome size estimate is considered robust, thus the assembled genome is estimated to cover 97% of the whole genome. A singe high 1C fluorescence peak for the *R. farinacea* sample suggests this individual was haploid.

Given that previously reported genome sizes for *Trebouxia* photobiont genomes estimated from genome sequencing assemblies are at least double the size of lichen mycobiont genomes (i.e., ca. 53–69 Mb; Beck et al. 2015; Kono et al. 2017; Tzovaras et al. 2020), and that no other peaks were detected in the flow cytometry analysis, we are confident that we were able to successfully separate the mycobiont tissue from the photobiont and that peak 1 in the flow histogram belongs to the mycobiont. These results demonstrate that mycobiont genome sizes can be accurately measured directly from lichen thalli, and that axenic culturing should no longer be a barrier for genome sizing of lichen mycobionts. This greatly improves the outlook for applying flow cytometry to lichen-genome-sequencing projects and supports previous work demonstrating the feasibility of flow cytometry for culture-recalcitrant fungal symbionts (Tavares et al. 2014). Our results also demonstrate that mycobiont genome size estimates from assembly size and k-mer profiles underestimate true genome size as measured by flow cytometry, a pattern recently observed in a Coleoptera data set (Pflug et al. 2020) and discussed by Hill et al. (2021) for fungi. Given the complex microbial communities found within lichen thalli and the risks associated with estimating assembly completeness using indirect metrics from assembly size, k-mer profiles, or BUSCO scores, we believe direct genome size measurement should be an essential part of future lichen metagenome genome sequencing projects.

Materials and Methods

Sample Collection

A fresh sample of *R. farinacea* was collected from a *Crataegus monogyna* subsp. *azarella* tree growing in the Royal Botanic Gardens Kew, UK. The thallus was cleaned of any substrate debris and 400 mg was weighed and transferred to a sterile 50 ml falcon tube. The sample was flash frozen in liquid nitrogen and ground to a fine powder using a chilled, sterile pestle and mortar. Ground material was transferred back to the falcon and immediately placed in a −80 °C freezer.

Flow Cytometry

Ramalina farinacea thalli are too small to use the same individual for both DNA extraction and flow cytometry. Therefore, further samples were collected from the same host tree at the same locality for flow cytometry. Nuclear DNA content was estimated by propidium iodide flow cytometry, and *C. micaceus* was used as the calibration standard using material from a single individual that had been isolated and cultured from a collection made by R. Wright 10/05/2020 at Royal Botanic Gardens Kew, UK (culture code: FTOL_0141). The genome size of *C. micaceus* was estimated by co-running a sample with *Arabidopsis thaliana* (L.) Heynh., 1842 (ecotype col-0 NASC) with an estimated genome size of 1C = 172.44 Mb. This was achieved by

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

using a new razor blade to co-chop a small amount of *C. micaceus* mycelium with 1 cm^2 fresh *A. thaliana* leaf tissue in a petri dish containing 1 ml of LB01 buffer (Doležel et al. 1989). A further 1 ml of LB01 was added to the sample and the contents gently mixed. The sample was then passed through a 30-µm nylon filter, stained with 100 µl propidium iodide (1 mg/ml) and incubated on ice for 15 min. Three replicates of the combined sample were run, recording up to 1,000 nuclei per fluorescence peak using a Sysmex CyFlow Space (Sysmex Europe GmbH, Norderstedt, Germany) flow cytometer fitted with a 100-mW green solid-state laser. The resulting histograms were analyzed with the Windows™-based FlowMax software (v. 2.9 2014; Sysmex Europe GmbH), and the average of each sample was used to estimate genome size (Pellicer et al. 2021).

Once the genome size of *C. micaceus* was estimated (1C = 62.62 Mb), it was then used as the internal calibration standard for estimating the genome size of *R. farinacea*, implementing the same method as above. Initial testing showed *C. micaceus* mycelium was rich in nuclei and therefore only one forceps pinch-worth was co-chopped with approximately 0.5 cm^2 of fresh *R. farinacea* thallus material.

DNA Extraction and Sequencing

DNA was extracted following the steps of a modified sodium dodecyl sulfate (SDS) extraction method (Merges et al. 2021), replacing SDS with cetyltrimethyl ammonium bromide (CTAB) buffer. See Supplementary Material online for a full DNA extraction protocol. The genomic DNA was eluted in 1 ml of TE buffer and column cleaned twice using QIAGEN mini-spin columns, before eluting into 50 µl of AE buffer from the Qiagen DNeasy® Plant Mini Kit (Qiagen, Redwood City, CA, USA). Extraction quality was checked using a Quantas™ Fluorometer (Promega, UK). DNA was sheared using a COVARIS-G tube, centrifuging at 4,200 rpm (1630 g) for 1 min to increase DNA throughput. The sheared DNA was converted into Oxford Nanopore libraries using the SQK-LSK110 library kit and the short fragment buffer (Oxford Nanopore Technologies, UK). Library prep followed the standard manufacturer's protocol apart from the final elution, which was done for 30 min at 37 °C. The libraries were sequenced using a SpotON R9.4.1 FLO-MIN106 flowcell for 48 h.

Basecalling and Metagenome Assembly

Raw sequences were basecalled using the super high accuracy model (SUP) of the GPU version of Guppy v6.0.1 (config file: dna_r9.4.1_450bps_sup.cfg; Oxford Nanopore Technologies Inc.). Basecalled reads were assembled into metagenome contigs using Flye v.2.9 with the "–meta" and "–nano-hq" options (Kolmogorov et al. 2019). The *k*-mer profile of unassembled mycobiont reads was computed using Jellyfish v2.2.6 ($k = 21$) (Marçais and Kingsford 2011). The resulting *k*-mer histogram was used as input for GenomeScope to estimate genome size (Vurture et al. 2017).

Mycobiont Filtering

Mycobiont reads were filtered from the metagenome assembly using the pipeline of Llewellyn et al. (2023). Briefly, the pipeline uses a BlobTools-based workflow (Laetsch and Blaxter 2017) and consists of DIAMOND (Buchfink et al. 2014) and BLAST+ blasts (Camacho et al. 2009), followed by reassembly of the filtered mycobiont reads. A detailed outline of all steps in each round of filtering can be found in the Supplementary Material online.

Terminal telomere sequences were identified by searching for the conserved fungal telomere sequence (CCCTAA/TTAGGG) using Tapestry v1.0. (https://github.com/johnomics/tapestry). Mycobiont genome assembly contiguity was assessed using QUAST v5.0.2 (Gurevich et al. 2013) and completeness assessed using BUSCO v4.0.2 (Manni et al. 2021) and the BUSCO *Ascomycota* lineage data set (ascomycota_odb10).

To confirm that the mycobiont assembly belonged to *R. farinacea*, we extracted the universal fungal barcode ITS sequence (ITS1, 5.8S, and ITS2) using ITSx (Bengtsson-Palme et al. 2013). ITS was blasted against the UNITE database to confirm its identity (Nilsson et al. 2019). We also used blastn to extract nucLSU, mitSSU, RPB1, and RPB2 genes with GenBank sequences JQ301589, KJ766480, KJ766831, and KJ766963, respectively, as queries. These four genes were used for phylogenetic placement using the Evolutionary Placement Algorithm as implemented in T-BAS v2.3 with the Lecanoromycetes v2 reference tree (see Supplementary Figure S1) (Miadlikowska et al. 2006; Carbone et al. 2017, 2019).

Annotation

A de novo repeat library was generated using RepeatModeler v2.0.1 with LTR discovery (Flynn et al. 2020) and used for softmasking with RepeatMasker v4.1.0 (Smit et al. 2015). Funnanotate was used to predict gene models (Palmer and Stajich 2019), incorporating expressed sequence tags and protein models from the JGI Mycocosm genomes of *Cladonia grayii*, *Usnea florida*, and *Xanthoria parietina* as evidence for EvidenceModeler (Haas et al. 2008).

Functional annotation of the predicted gene models was performed using InterProScan v5.57-90.0 (Jones et al. 2014); eggNOG-mapper v2.1.9-4dfcbd5 (Cantalapiedra et al. 2021) with sequences searched against the eggNOG orthology database v5.0.2 (Huerta-Cepas et al.

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

2019) using DIAMOND v2.0.15 (Buchfink et al. 2014); and antiSMASH v6.1.1 (Blin et al. 2021).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Dr Francesco Dal Grande and Dr Jessica Allen for their advice on DNA extraction methods and Dr Alexandra Dallaire for her advice on Nanopore library preparation. The authors acknowledge the assistance of the ITS Research team at Queen Mary University of London. This work was funded by the Natural Environment Research Council (NERC) studentship NE/S007415/1 and the Emily Holmes Scholarship fund to T.L.

Data Availability

WGS data and the annotated genome assembly are available on GenBank under the BioProject accession PRJNA890456. All bioinformatics scripts are available at https://github.com/theo-llewellyn/longread-lichens.

Literature Cited

Allen JL, Jones SJM, McMullin RT. 2021. Draft genome sequence of the lichenized fungus *Bacidia gigantensis*. Microbiol Resour Announc. 10:e00686-21.

Beck A, Divakar PK, Zhang N, et al. 2015. Evidence of ancient horizontal gene transfer between fungi and the terrestrial alga Trebouxia. Org Divers Evol. 15:235–248. https://doi.org/10.1007/s13127-014-0199-x

Bengtsson-Palme J, et al. 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol. 4:914–919.

Blin K, et al. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49:W29–W35.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12:59–60.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38:5825–5829.

Carbone I, et al. 2017. T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. Bioinformatics 33:1160–1168.

Carbone I, et al. 2019. T-BAS version 2.1: Tree-Based Alignment Selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. Microbiol Resour Announc. 8:e00328-19.

Doležel J, Binarová P, Lcretti S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. Biol Plant. 31:113–120.

Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117:9451–9457.

Gerasimova JV, Beck A, Werth S, Resl P. 2022. High diversity of type I polyketide genes in *Bacidia rubella* as revealed by the comparative analysis of 23 lichen genomes. J Fungi. 8:449.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9:R7.

Hill R, Leitch IJ, Gaya E. 2021. Targeting Ascomycota genomes: what and how big? Fungal Biol Rev. 36:52–59.

Huerta-Cepas J, Szklarczyk D, Heller D. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47:D309–D314. https://doi.org/10.1093/nar/gky1085

Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37:540–546.

Kono M, Tanabe H, Ohmura Y, Satta Y, Terai Y. 2017. Physical contact and carbon transfer between a lichen-forming Trebouxia alga and a novel Alphaproteobacterium. Microbiology 163:678–691. https://doi.org/10.1099/mic.0.000461

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Research 6:1287.

Le Cam B, et al. 2019. Population genome sequencing of the scab fungal species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata*. G3 Genes|Genomes|Genetics. 9:2405–2414.

Llewellyn T, et al. 2023. Metagenomics shines light on the evolution of "sunscreen" pigment metabolism in the Teloschistales (lichen-forming Ascomycota). Genome Biol Evol. 15:evad002.

Lücking R, Hodkinson BP, Leavitt SD. 2017. The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota—approaching one thousand genera. Bryologist 119:361–416.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 38:4647–4654.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770.

McKenzie SK, Walston RF, Allen JL. 2020. Complete, high-quality genomes from long-read metagenomic sequencing of two wolf lichen thalli reveals enigmatic genome architecture. Genomics 112:3150–3156.

Merges D, Dal Grande F, Greve C, Otte J, Schmitt I. 2021. Virus diversity in metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes, putative hosts and elevational distributions. Environ Microbiol. 23:6637–6650.

Miadlikowska J, et al. 2006. New insights into classification and evolution of the Lecanoromycetes (Pezizomycotina. Ascomycota) from phylogenetic analyses of three ribosomal RNA- and two protein-coding genes. Mycologia 98:1088–1103.

Miyauchi S, et al. 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat Commun. 11:5125.

Molins A, Moya P, Muggia L, Barreno E. 2021. Thallus growth stage and geographic origin shape microalgal diversity in *Ramalina farinacea* lichen holobionts. J Phycol. 57:975–987.

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023

Nilsson RH, et al. 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. 47:D259–D264.
Palmer J, Stajich J. 2019. nextgenusfs/funannotate: funannotate v1.5.3. https://zenodo.org/record/2604804
Peay KG, Kennedy PG, Talbot JM. 2016. Dimensions of biodiversity in the Earth mycobiome. Nat Rev Microbiol. 14:434–447.
Pellicer J, Powell RF, Leitch IJ. 2021. The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants. In: Besse P, editor. Molecular plant taxonomy: methods and protocols. Methods in molecular biology. New York (NY): Springer. p. 325–361.
Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. 2020. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3 (Bethesda) 10:3047–3060.
Singh G, Armaleo D, Dal Grande F, Schmitt I. 2021. Depside and depsidone synthesis in lichenized fungi comes into focus through a genome-wide comparison of the olivetoric acid and physodic acid chemotypes of *Pseudevernia furfuracea*. Biomolecules 11: 1445.
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. http://repeatmasker.org
Stocker-Wörgötter E, Elix JA, Grube M. 2004. Secondary chemistry of lichen-forming fungi: chemosyndromic variation and DNA-analyses of cultures and chemotypes in the *Ramalina farinacea* complex. Bryo 107:152–162.
Tagirdzhanova G, et al. 2021. Predicted input of uncultured fungal symbionts to a lichen symbiosis from metagenome-assembled genomes. Genome Biol Evol. 13:evab047.
Tavares S, et al. 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. Front Plant Sci. 5:422.
Tzovaras BG, et al. 2020. What is in *Umbilicaria pustulata*? A metagenomic approach to reconstruct the holo-genome of a lichen. Genome Biol Evol. 12:309–324.
Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204.
Wilken PM, et al. 2020. IMA genome - F13 draft genome sequences of *Ambrosiella cleistominuta*, *Cercospora brassicicola*, *C. citrullina*, *Physcia stellaris*, and *Teratosphaeria pseudoeucalypti*. IMA Fungus 11:19.

Associate editor: Li-Jun Ma

Downloaded from https://academic.oup.com/gbe/article/15/5/evad074/7156542 by guest on 07 July 2023