Genome-wide analyses of biosynthetic genes in lichen-forming fungi

Dissertation

zu Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Anjuli Calchera

aus Frankfurt am Main

Frankfurt (2019)

(D 30)

vom Fachbereich Biowissenschaften der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel
Institut für Ökologie, Evolution und Diversität
Johann Wolfgang Goethe - Universität
D-60438 Frankfurt am Main

Gutachter: Prof. Dr. Imke Schmitt
Institut für Ökologie, Evolution und Diversität
Johann Wolfgang Goethe - Universität
D-60438 Frankfurt am Main

Prof. Dr. Markus Pfenninger
Institut für Organismische und Molekulare Evolutionsbiologie
Johannes Gutenberg - Universität Mainz
D-55128 Mainz

Datum der Disputation: 24.06.2020

This thesis is based on the following publications:

Meiser, A., Otte, J., Schmitt, I., & Dal Grande, F. (2017). Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi. *Scientific Reports*, 7(1), 14881, doi:10.1038/s41598-017-14576-6.

Dal Grande, F., Meiser, A., Greshake Tzovaras, B., Otte, J., Ebersberger, I., & Schmitt, I. (2018a). The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *The Lichenologist*, 50(3), 329–340, doi:10.1017/S002428291800021X.

Calchera, A., Dal Grande, F., Bode, H. B., & Schmitt, I. (2019). Biosynthetic gene content of the 'perfume lichens' *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules*, 24(1), 203, doi:10.3390/molecules24010203.

Contents

List of Figures

List of Tables

1. Abstract

In the light of emerging resistances against common drugs, new drug leads are required. In the past natural sources have been more yielding in this respect than synthetic strategies. Fungi synthesize many natural products with biological activities and pharmacological relevance. However, only a fraction of the estimated fungal diversity has been evaluated for biological activity, and much of the Fungi's natural chemical diversity awaits discovery. Especially promising in this context are lichenized fungi. Lichens are well known for their particularly rich and characteristic secondary chemistry which allows them to withstand intense UV radiation, protects them against herbivory, and prevents them from being overgrown. The slow growth rates of lichens and difficulties and infeasibility of large scale cultivations in the laboratory render lichens inaccessible for applied purposes. These experimental challenges have led to a poor understanding of the molecular mechanisms underlying the biosynthesis of characteristic lichen secondary metabolites. The recent development of improved sequencing techniques has enabled new strategies to address multi-species assemblages directly through metagenome sequencing and survey their biosynthetic potential through genome mining. However, whole genome sequencing of entire lichen thalli to metagenomically assess the lichen-forming fungus without the need of cultivation has not been evaluated for lichens before. This approach will enable the reconstruction of fungal genomes from mixed DNA from lichen thalli and allow the exploration of biosynthetic gene content.

My thesis was conducted in two parts: a methodological evaluation of a metagenomic strategy to reconstruct genomes and gene sets of lichen-forming fungi, and the exploration of biosynthetic gene content with the help of comparative genomics and phylogenetics. For the first part, I evaluated the quality of metagenome-derived genome assemblies and gene sets by direct comparison to culture-derived reference assemblies and gene sets of the same species. I showed that metagenome-derived fungal assemblies are comparable to culture-derived references genomes and have a similar total genome size and fungal genome completeness. The quality of assemblies was affected strongly by the choice of assembler, but not by the method of taxonomic assignment or

inference of non-mycobiont DNA sequences. The fungal gene space is well covered in metagenome-derived and culture-derived fungal gene sets and overlaps to 88-90 %. Finally, the metagenome-derived assemblies reliably recover gene families of secondary metabolism. This shows the suitability of metagenomically derived genomes for mining biosynthetic genes, and potentially also other gene families. Overall, the method validation showed a high similarity between metagenome- and culture-derived genome assemblies.

For the second part of my thesis, I explored the biosynthetic gene content in two different systems: Between two sister-species with different ecological requirements but similar chemical profile, and between two species which are metabolite-rich and economically relevant in the perfume industry. I compared the diversity of biosynthetic gene clusters between the species and in the broader context of other lichenized and non-lichenized fungi. Overall, the whole genome mining revealed a large number of uncharacterised secondary metabolite gene clusters in fifteen genomes of lichen-forming fungi compared to other fungal classes. Their number highly outweighs the number of known synthesized metabolites and highlights the hidden biosynthetic potential in lichen-forming fungi. Many biosynthetic gene clusters in the ecological distinct sister-species showed a high homology in accordance with the high synteny in gene content and order in both genomes. These clusters represent ideal candidates for secondary metabolites synthesized by both species, while the remaining clusters may encode for metabolites relevant for the different ecological requirements of both species. The metabolite-rich species used in the perfume industry showed a particularly high number of biosynthetic gene clusters. An in-depth characterization of architecture and gene content of homologous gene clusters together with hints from phylogenetic relatedness to functional characterized metabolites provides promising insights into the biosynthetic gene content of these lichen-forming fungi.

In conclusion, I showed that metagenome sequencing of natural lichen thalli is a feasible approach to reconstruct the fungal mycobiont genome of lichens and circumvent time-consuming and in some cases impossible cultivation of individuals. The genome mining for secondary metabolite gene clusters in lichen-forming fungi revealed a high biosynthetic potential for the discovery of new natural products. One of the focal species, *Evernia prunastri*, contained the highest ever reported number (80) of biosynthetic clusters in lichenized fungi. The comprehensive cluster characterizations through annotation, comparative mapping and phylogenetics provide first valuable hints for linking metabolites to genes in these lichen-forming fungi. My results pave the way for

biotechnological strategies to unlock the vast richness of natural products from lichens for applied purposes.

2. Introduction

2.1. Natural products from fungi

Fungi synthesize a large array of natural products, many of which show biological activities. Natural products are secondary metabolites, which play important roles in the modulation of biological systems and organism - environment interactions, including warfare, defence and development (Keller, 2019; Mouncey *et al.*, 2019). Primary metabolites, in contrast, are essential for growth, development and reproduction of an organism (Keller, 2019). Some natural products from fungi are potent toxins, such as aflatoxin, which can cause severe food poising (Wogan, 1966), some have industrial and medical importance, for example as antibiotics (penicillin), immunosuppressive agents (cyclosporine), or cholesterol lowering drugs (lovastatin) (Brakhage, 2013).

Most potential drugs leads have been inspired by, and generated from natural sources (Mishra & Tiwari, 2011). Only 27 % of all newly approved drugs (1981-2014) are totally synthetic (Newman & Cragg, 2016, see Fig. 2.1). Natural products are a rich source of compounds for drug discovery and they form, together with their derivates and natural product-inspired synthetic compounds, the backbone of most of our current drugs in use (Wright, 2019). In addition, the biologically relevant chemical space is better covered by natural products than by synthetic compounds (Harvey *et al.*, 2015). There is a growing need for the discovery of new compounds that may prove

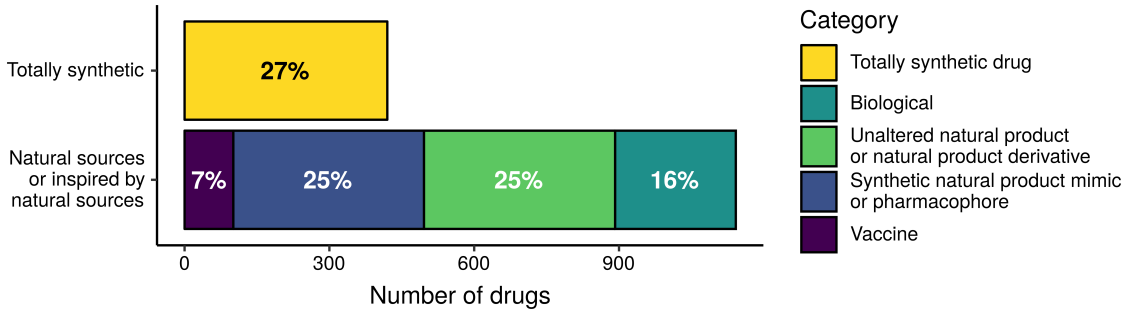


Figure 2.1. The majority of newly approved drugs in the years between 1981 and 2014 was either produced from a natural source or inspired by a natural source. Modified after Newman & Cragg (2016).

efficient as drugs against the emergence of multidrug-resistant pathogenic strains and increasing resistances against common antibiotics (Bode & Müller, 2005; Lewis, 2013). Natural products and natural product derived compounds are often more complex and therefore more effective against increasing antibiotic resistance.

Since less than 10 % of world's biodiversity has been screened for biological activity, natural sources remain to be a promising source for the discovery of new lead compounds (Cragg & Newman, 2005; Dias *et al.*, 2012). Novel natural product skeletons, which can be used as leads for pharmaceuticals are likely to come from genomic sources in the future (Newman & Cragg, 2016). The new genomic era provides the tools to circumvent the need to isolate and culture individual organisms to access natural product biosynthetic genes (Wright, 2019). Besides well studied and accessible organisms a rich source of novel compounds is likely to come from non-traditional sources (Harvey *et al.*, 2015). For example, symbiotic organisms have been untapped in a systematic way as source of natural products (Wright, 2019). Among fungi, symbiotic lichenized fungi are predicted to hold a particular rich secondary metabolome space (Bills & Gloer, 2016; Keller, 2019).

2.2. Natural products from lichens

The secondary chemistry of lichen-forming fungi is particularly rich, and has been investigated comprehensively for over a century (Asahina & Shibata, 1954; Culberson, 1969; Elix, 2014; Elix *et al.*, 1984; Huneck & Yoshimura, 1996; Zopf, 1907). Chemistry plays an important role in the classification of lichens (chemotaxonomy; Lumbsch, 1998). Some species of lichenized fungi are difficult to differentiate morphologically, but can be distinguished by their often considerable quantities of secondary metabolites (Huneck & Yoshimura, 1996). Lichens are commonly described as symbiotic associations of a fungal partner (mycobiont), and one or more photosynthetic partners (photobiont) (Nash, 2008). Ninety-eight percent of lichen-forming fungi are Ascomycetes and the photobiont is typically either an eukaryotic green alga (e.g. *Trebouxia*, *Trentepohlia*) or a prokaryotic cyanobacterium (e.g. *Nostoc*) or both. Within Fungi, lichenization is one of the most common lifestyles, as almost one fifth of all described fungi are lichenized ($\sim$ 19,000 species) (Lücking *et al.*, 2016). The lichen symbiosis is very successful and can be found worldwide in most terrestrial ecosystems from polar regions to the tropics. As lichens, both mycobionts and photobionts have expanded into many habitats, that would have been unsuitable as free-living organisms (Nash, 2008).

The traditional view of lichens as a dual symbiosis of a dominating primary fungal symbiont with its characteristic photosynthetic partner has been recently challenged by microbiome studies showing that lichens indeed are complex multi-species symbiotic assemblages that form microhabitats harbouring a high diversity of other eukaryotic and prokaryotic microorganisms (Aschenbrenner *et al.*, 2016; Cernava *et al.*, 2017; Grube & Wedin, 2016). The advent of new high-throughput sequencing technologies has opened opportunities for studying the presence of multiple fungal and algal species together with rich bacterial communities within individual lichen thalli (Dal Grande *et al.*, 2018b; Fernández-Mendoza *et al.*, 2017; Machado *et al.*, 2017; Moya *et al.*, 2017; Muggia & Grube, 2018). Furthermore it has been proposed that one or in some cases even two specific basidiomycete yeasts are commonly associated as lichen-inhabiting fungi in the cortex of several widespread lichen genera (Spribille, 2018; Spribille *et al.*, 2016; Tuovinen *et al.*, 2019).

Lichens are known to impact several community processes, especially through their rich and diverse secondary chemistry (Asplund & Wardle, 2017; Seaward, 2008). More than 1,000 secondary metabolites are known from lichens and many are exclusively found in lichens (Elix & Stocker-Wörgötter, 2008; Huneck & Yoshimura, 1996; Molnár & Farkas, 2010; Stocker-Wörgötter, 2008). The vast majority of these characteristic lichen secondary metabolites is of fungal origin and most commonly deposited extracellularly, e.g. on the outer surface of medullary hyphae (Elix & Stocker-Wörgötter, 2008). Lichen substances are typically unevenly distributed throughout the thallus: pigments are often restricted to the upper cortex, while colourless substances can be restricted to the medullary layer (Elix & Stocker-Wörgötter, 2008). The concentrations of lichen substances vary typically between 0.1 % and 5 % of the lichen thallus dry weight, but can sometimes account for up to 40 % (Rundel, 1978; Stocker-Wörgötter, 2008). Secondary metabolites in lichens have various ecological roles including light-screening, chemical weathering, allelopathic and anti-herbivore defence (Huneck, 1999; Lawrey, 1986; Rundel, 1978). Besides the benefits for the lichens themselves, the compounds have manifold biological activities, some of which are of pharmaceutically interest. Activities of lichen compounds include antiviral, antibiotic, antitumor, allergenic, plant growth inhibitory, antiherbivore, and enzyme inhibitory activities (Boustie & Grube, 2005; Huneck, 1999; Molnár & Farkas, 2010; Ranković & Kosanić, 2015). For example, the natural products evernic acid, usnic acid, physodic acid and atranorin produced by the lichens *Evernia prunastri* and *Pseudevernia furfuracea* are strong metabolic enzyme inhibitors. It has been reported, for example that atranorin may inhibit tumorigenesis

and lung cancer cell mobility (Boustie & Grube, 2005; Zhou *et al.*, 2017). Gyrophoric acid, known as major compound from *Umbilicaria pustulata* and *Umbilicaria hispanica*, is a potent antiproliferative agent against the growth of human keratinocytes (Kumar KC & Müller, 1999), thus representing an interesting lead in the promotion of tissue regeneration (Burlando *et al.*, 2009). A wide range of lichen genera, including *Evernia*, *Pseudevernia* and *Umbilicaria*, have already been used in traditional medicine all around the world, for example to treat wounds, skin disorders, respiratory, digestive, obstetric and gynecological issues (Crawford, 2015). Lichen natural products therefore are promising targets for further development of pharmaceutical applications.

However, two main reasons hinder the mainstream use of lichen compounds i) the low growth rate of many lichen species, ii) the difficulty or, in some cases, impossibility of isolating and *in vitro* cultivating the symbionts (Crittenden *et al.*, 1995; Grube *et al.*, 2013; McDonald *et al.*, 2013a). Harvesting and isolating compounds of interest directly from lichen thalli is usually not sustainable and often not feasible, due to the large amount of biomass needed (Miao *et al.*, 2001). The difficulties in aposymbiotic mycobiont cultivation and the slow growth rates in lichen-forming fungi pose particular challenges for the experimental characterization of the molecular mechanisms underlying the biosynthesis of characteristic lichen metabolites (Abdel-Hameed *et al.*, 2016b; Stocker-Wörgötter, 2008). The biosynthesis of natural products may also be influenced by abiotic factors such as microhabitat, elevation or the response to hydration, chemical signals, UV radiation, seasonality or microclimatic fluctuations (Boustie & Grube, 2005; Stocker-Wörgötter, 2008). The biosynthesis of substances may also depend on biological factors, such as contact to microorganisms within the lichen thallus, predation by insects or snails, or the presence of competing organisms (Calcott *et al.*, 2018). Furthermore, mycobionts synthesize significant amounts of natural products only under specific triggering conditions and therefore the secondary metabolite production in axenic cultures can differ substantially from that found in nature (Boustie & Grube, 2005; Muggia *et al.*, 2009; Ranković & Kosanić, 2015).

2.3. Lichen genomics

Genome sequences of obligate symbionts can be obtained by accessing individual symbionts directly, or by retrieving the information from multi-species, metagenomic samples (Grube & Wedin, 2016). The traditional method to acquire genetic information

is to obtain pure, single-species DNAs through the separation and cultivation of individual symbionts. However, in obligate symbioses, this poses the challenge of aposymbiotic cultivation of each symbiont. The advance of high-throughput sequencing technology has now made possible the study of slow growing symbiotic organisms without the need for establishing and growing axenic cultures of the individual symbionts (Grube *et al.*, 2013). Using a metagenomic shotgun sequencing approach, it is possible to reconstruct high-quality genomes from multi-species samples. In this context, one option is the DNA enrichment for targeted organisms by removing unwanted tissue or cells before or during DNA extraction, increasing the targeted cell portion by amplification, or the application of single-cell sequencing. Examples utilizing some form of DNA enrichment include studies targeting specific bacterial groups associated with a marine sponge (Wilson *et al.*, 2014), the obligate fungal symbiont of aphids (Vogel & Moran, 2013), or the primary fungal symbiont of lichens (McDonald *et al.*, 2013b). Another metagenomic approach is to sequence the total symbiotic association as metagenomic sample first, only subsequently assigning the genetic information to the symbiotic partners via bioinformatic analysis (Grube & Wedin, 2016).

The method of reconstructing genome sequences directly from metagenomic samples circumvents experimentally tedious cultivation or enrichment procedures, but can also be methodologically demanding (Breitwieser *et al.*, 2019; Greshake *et al.*, 2016; Tringe & Rubin, 2005). The proportion of available genetic information can be shifted towards one symbiont in the association (e.g. the host species or primary symbiont), leading to highly uneven sequencing depth and genome coverages of individual partners (Grube *et al.*, 2013; Peng *et al.*, 2012; Wooley *et al.*, 2010). Some assemblers, including dedicated tools for metagenomic samples, can be surprisingly sensitive to low and uneven coverages (Greshake *et al.*, 2016). Hence, the quality of genome reconstruction depends essentially on the choice of assembler. Furthermore, challenges arise from the risk of generating chimeric contigs, which are assembled from sequences originating from multiple genomes (Nurk *et al.*, 2017). Especially repetitive DNA regions of single genomes, homologous regions of closely related strains, or conserved regions of different species are difficult to assemble and failure to resolve these regions may result in rearrangement errors and intergenomic assembly errors (Sangwan *et al.*, 2016). Further challenges include post-assembly sequence binning procedures, where contigs are taxonomically assigned to phylogenetic groups. Features that are used to bin contigs to taxonomic groups include first of all similarity to sequenced genomes but also GC content, oligonucleotide frequency, and coverage depth (Tringe & Rubin,

2005). Inaccurate assignments can result into chimeric genome bins, i.e. groups of contigs from multiple species. Additionally, the assignment success of a similarity-based classification relies on the availability of high-quality, accurate, and contamination-free reference databases, especially for whole genome analysis (Breitwieser *et al.*, 2019).

Genomes of lichen-forming fungi have become publicly available since 2011 (Grube *et al.*, 2013; Grube & Wedin, 2016; Werth *et al.*, 2015). Most available genomes still originate from cultivated mycobionts (e.g. Abdel-Hameed *et al.*, 2016b; Armaleo *et al.*, 2019; Armstrong *et al.*, 2018; Park *et al.*, 2013a,b, 2014a,b,c; Wang *et al.*, 2018b, 2014b). Given the difficulties of axenic cultivation, metagenomic approaches represent particularly relevant tools to access the genomic features of lichen-forming fungi, in particular for species that are difficult to isolate and grow *in vitro*. Therefore researchers have begun to implement metagenomics tools to study ecological, evolutionary and biotechnological aspects of the lichen symbiosis. Examples applying high-throughput metagenome or metatranscriptome sequencing directly from lichen thalli include mitochondrial genome assessments of *Peltigera membranacea* and *Peltigera malacea* (Xavier *et al.*, 2012), gene expression studies in *P. membranacea* (Miao *et al.*, 2012), transcriptome characterization of *Cladonia rangiferina* (Junttila & Rudd, 2012), secondary metabolite biosynthetic pathway discovery in *P. membranacea* (Kampa *et al.*, 2013), ammonium transporter gene detection in several lichen species (McDonald *et al.*, 2013b), functional endosymbionts analyses in *Lobaria pulmonaria* (Erlacher *et al.*, 2015) or genetic population structure (Allen *et al.*, 2018). Altogether, high-throughput metagenome sequencing constitutes an indispensable and valuable tool to study especially uncultivable lichen-forming fungi and circumvent cultivation challenges. However, to date only a handful of studies has implemented approaches to reconstruct genomes of lichen-forming fungi from whole metagenomic lichen thalli. The accuracy and completeness of metagenomically assembled genomes of lichenized fungi remain to be evaluated.

2.4. Genome mining of secondary metabolite gene clusters

Despite the rich chemical diversity found in lichens, biotechnological applications remain limited since only a few genomes and biosynthetic studies of lichen-forming fungi are available. Although chemically diverse, all secondary metabolites are produced by a few common biosynthetic pathways (Keller *et al.*, 2005). The diverse range of lichen substances can be classified into three major groups according to their biosynthesis

in the mycobiont (Huneck, 1999). The majority of secondary metabolites present in lichens are derived from the acetyl-polymalonyl pathway (polyketide pathway) and includes the substance classes of depsides, depsidones, dibenzofurans, anthraquinones, chromones, and xanthones (Calcott *et al.*, 2018; Elix & Stocker-Wörgötter, 2008). Also common are metabolite classes derived from the shikimic acid pathway, for example terphenylquinones and pulvinic acid derivates, and metabolite classes derived from the mevalonic acid pathway, for example terpenes, steroids and carotenoids (Elix & Stocker-Wörgötter, 2008; Goga *et al.*, 2018).

The most common classes of lichen secondary metabolites are built by polyketide synthase (PKS) multienzymes through the acetyl-polymanoyl pathway. (Boustie & Grube, 2005; Nash, 2008). These enzymes link derived monoaromatic subunits (orcinol, β-orcinol type or methylphloroacetophenone) by ester, ether or carbon-carbon bonds to form phenolic compounds, dibenzofurans, depsides and depsidones (Calcott *et al.*, 2018). Fungal PKSs are multidomain systems, that reuse a set of active site domains iteratively through multiple catalytic cycles to elongate their polyketide products (Cox *et al.*, 2018). PKSs are grouped according to the level of reductive processing into partially or highly reducing polyketide synthase (R-PKS) and non-reducing polyketide synthase (NR-PKS) (Crawford & Townsend, 2010), and both types of polyketides can be found in lichen-forming fungi. However, the most common lichen substance classes such as depsides and depsidones are derived from NR-PKSs (Elix & Stocker-Wörgötter, 2008; Schmitt *et al.*, 2005). Another class of secondary metabolites are non-ribosomal peptides produced by multidomain, multimodular enzymes named non-ribosomal peptide synthetases (NRPSs). NRPSs assemble amino acids into peptides through amide bonds and are independent of the ribosomal machinery. Furthermore, the PKS and NRPS pathways can be combined to form hybrid polyketide synthase – non-ribosomal peptide synthetases (PKS-NRPSs) that build structurally complex hybrid molecules fusing polyketides and amino acids by an amide bond (Boettger & Hertweck, 2013). The typical domain architecture of PKSs and NRPSs as common classes of secondary metabolite gene families are shown in Fig. 2.2.

Genes encoding specific fungal secondary metabolites are often arranged in a contiguous cluster (Keller *et al.*, 2005). A typical fungal secondary metabolic gene cluster is composed of a chemically defining backbone or core synthase or synthetase gene (e.g. PKS, NRPS, terpene synthase) and genes encoding tailoring functions that can modify the carbon backbone chemically (Keller, 2019; Rokas *et al.*, 2018). Often transporter genes and pathway-specific regulatory genes are also part of the metabolic gene clus-

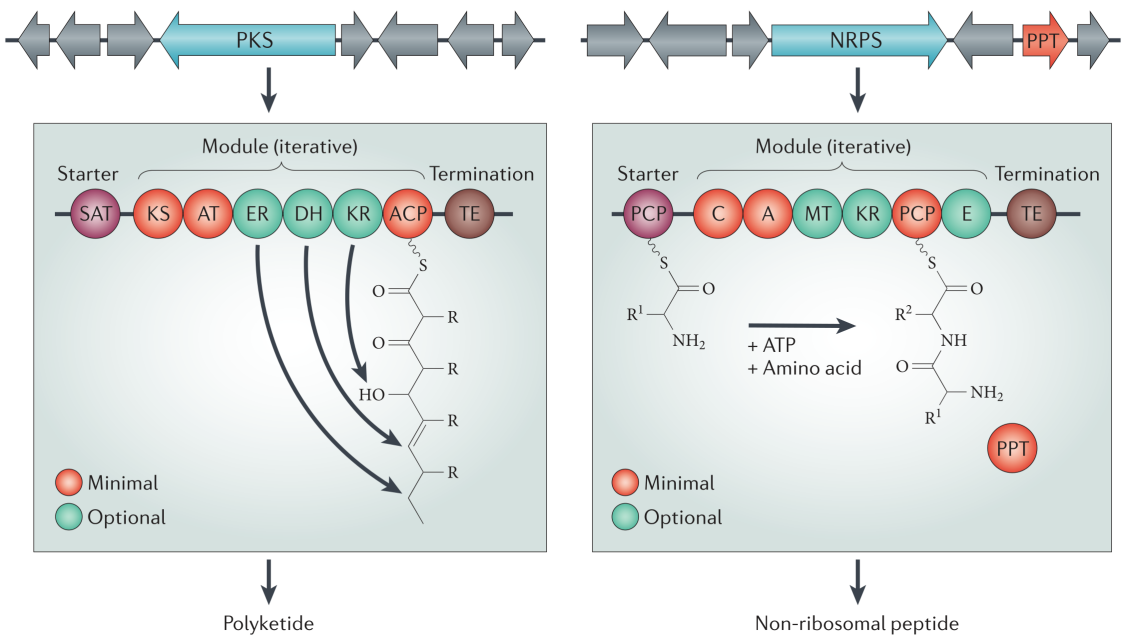


Figure 2.2. Common classes of fungal secondary metabolite gene clusters. Typical protein domain architecture of polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) genes are shown with their catalytic principle. For abbreviations see the Glossary on page 47. Figure from (Brakhage, 2013).

ter. The fact that genes involved in secondary metabolite biosynthesis are clustered together in the genome can be exploited by advanced bioinformatic mining algorithms (Nielsen *et al.*, 2017). Recent improvements in genome sequencing techniques have drastically increased the number of available genomes that can be searched by these computational mining tools and have promoted natural product research. Consequently, genome mining has lead to the discovery of thousands of biosynthetic gene clusters (Cimermancic *et al.*, 2014; Doroghazi *et al.*, 2014; Zhao & Medema, 2016). However, most of these identified gene clusters encode metabolites of unknown structure and function and are therefore often referred to as orphan or cryptic gene clusters (Blin *et al.*, 2017; Brakhage, 2013). Furthermore, the high number of biosynthetic gene clusters predicted via genome sequencing often does not correspond to the known expressed metabolic profile of the species, thus hinting at a plethora of undiscovered chemical compounds (Bergmann *et al.*, 2007; Rokas *et al.*, 2018). For example, Nielsen *et al.* (2017) discovered almost 800 biosynthetic gene clusters containing a PKS, NRPS or hybrid PKS-NRPS in 24 *Penicillium* genomes, but were only able to connect a small fraction to known compounds from those species. In another example, Bertrand *et al.* (2018a) found a rich genetic secondary metabolic profile consisting of several orphan

gene clusters in the genome of a mycobiont of *Cladonia uncialis* known to produce only a single compound, i.e. usnic acid.

Functional characterization of fungal PKS gene clusters is particularly challenging. In contrast to modular bacterial PKSs, fungal PKSs function in an iterative fashion that conceals a cryptic programming in the PKS itself and makes the polyketide structure unpredictable (Keller *et al.*, 2005). Fungal PKSs can catalyse different sets of reactions during each extension cycle using the same catalytic domains and can generate diverse and complex molecules starting from simple building blocks (Hertweck, 2009). Our understanding of this intrinsic programming mechanisms that control starter unit selection, chain length, and the degree of reduction, dehydration and methylation in iterative PKS is still limited (Cox *et al.*, 2018; Schümann & Hertweck, 2006). Furthermore, the specificity and activity of tailoring enzymes are mostly unknown and further hamper structure prediction (Cox *et al.*, 2018). Experimental characterization aimed at connecting genes to metabolite production – for example through heterologous expression or gene knockout – have been performed only on a fraction of the many biosynthetic gene clusters predicted in sequenced genomes (Kautsar *et al.*, 2019).

Functional characterization of secondary metabolites in lichen-forming fungi is hampered by slow growth rates and difficulties in establishing pure cultures needed for direct experimental genetic manipulations (Gagunashvili *et al.*, 2009). Consequently no secondary metabolite has been directly linked to its encoding biosynthetic gene clusters with experimental proof (Bertrand & Sorensen, 2019a). However, several putative assignments have been proposed through a variety of methods. For the detection of PKS genes in lichenized fungi various methods have been established including the use of degenerate primers to amplify PKS gene fragments from genomic DNA and designing probes for library screening, and the use of complementary DNA (cDNA)-based templates generated from reverse-transcribed messenger RNA (mRNA) (Bertrand & Sorensen, 2018). Since the first transcriptomes and genomes of lichen-forming fungi are becoming available, biosynthetic insights can also be gained directly from the sequencing for mycobiont cultures or reconstructed from metagenomic thalli (Bertrand *et al.*, 2018b; Wang *et al.*, 2018a). The function of PKS genes can be predicted by experimental evidence such as gene knockout, heterologous expression (Chooi *et al.*, 2008; Gagunashvili *et al.*, 2009), radiolabeling (Taguchi *et al.*, 1969) or oxidation experiments (Barton *et al.*, 1956). More recently new methods for inferring functions have included phylogenetic relatedness to characterized genes (Wang *et al.*, 2014a; Ziemert & Jensen, 2012), transcriptional profiling (Armaleo *et al.*, 2011; Wang *et al.*,

2018a) or comparative homology mapping of entire gene clusters (Bertrand *et al.*, 2018b; Cox *et al.*, 2018; Nielsen *et al.*, 2017). Often several approaches are combined to provide putative assignments (Wang *et al.*, 2018a). The first putative assignment of a lichen biosynthetic gene and its corresponding secondary metabolite, the lichen-specific depsidone grayanic acid, was achieved by Armaleo *et al.* (2011) in *Cladonia grayi*. The link was proposed based on a combination of phylogenetic analysis, biosynthetic predictions from PKS domain arrangement, and correlation of mRNA levels with metabolite production. Later, the gene clusters encoding usnic acid and 6-hydroxymellein have been putatively identified in *Cladonia uncials* together with the putative functional prediction of tailoring enzymes encoded in the same clusters for prediction (Abdel-Hameed *et al.*, 2016a,b). However, all attempts to heterologously express lichen-derived PKS genes in a cultivable host have failed. No products were ever detected, even when a number of accessory cluster genes were successfully co-expressed (Armaleo *et al.*, 2011; Bertrand & Sorensen, 2019a,b; Chooi *et al.*, 2008; Gagunashvili *et al.*, 2009; Wang *et al.*, 2016). A review by Bertrand & Sorensen (2018) provides an overview of PKS genes identified from several species of lichenized fungi and summarizes methods used for detecting and characterizing lichen PKS genes and highlights the prospects of genomics-driven natural product discovery in lichens.

2.5. Knowledge gap

Lichens produce a vast diversity of secondary metabolites with a variety of biological activities and therefore represent a promising source for pharmaceutically interesting natural products. However, several aspects hamper tapping lichens as valuable new source in natural product research. No definite link between a lichen secondary metabolite and the encoding biosynthetic gene cluster has been made to date. The main reason for this missing link lies in the fact that lichens are experimentally challenging. Lichens are slow growing symbiotic multi-species communities which cannot be harvested sustainably. The cultivation of mycobionts as source of the metabolites of interest is time-consuming and often not feasible. Furthermore, the biosynthesis of lichen natural products may depend on many abiotic or biotic factors that are experimentally hard to reproduce. A promising alternative to circumvent cultivation has recently arisen with the advent of next generation sequencing techniques. Hereby the individual genomes of the symbiotic partners have to be reconstructed from complex DNA mixtures, an approach that is still in its infancy in eukaryotes and has not been thoroughly evaluated

for lichen symbioses. Both options for obtaining genomes of lichen-forming fungi – mycobiont cultivation or metagenomic reconstruction – remain challenging, and only a few genomes have become available during the last decade. However, high quality genomes are necessary to mine biosynthetic genes from lichenized fungi, and to unlock these organisms for biotechnological applications.

3. Thesis aims and structure

The overall aim of the present thesis is the implementation of metagenomic approaches to better understand the biosynthetic gene content of lichen-forming fungi. This provides a basis for linking genes to secondary metabolites, and to unlock the biosynthetic potential of lichens for biotechnological purposes.

Specifically, I first evaluated the quality of metagenomically assembled genomes of lichen-forming fungi (**Q1**). Second, I investigated the biosynthetic gene content of two closely related species (**Q2**) and two metabolite-rich species (**Q3**). I make use of comparative genomics and phylogenetics to bridge the missing link between biosynthetic gene cluster and secondary metabolite richness.

Q1: How well can we reconstruct the genomes and gene content of lichen-forming fungi from metagenomic samples?
In this study I compare the genomes and gene space of two lichen-forming fungi, each retrieved from a) a complex metagenomic DNA mixture from a lichen thallus and b) an axenically cultivated mycobiont. I also access to what extent contamination by non-mycobiont DNA affects the ability to reconstruct genomes of lichenized fungi from metagenomic samples. Results of this study are published in Meiser *et al.* (2017, Appendix A.1).

Q2: What is the biosynthetic content of two sister-species with different ecological requirements?
In this study we generate a *de novo* genome of *Umbilicaria hispanica* and compare it to the closely related *Umbilicaria pustulata*. Specifically we assess gene homology between the species, and differences in biosynthetic gene clusters. Results of this study are published in Dal Grande *et al.* (2018a, Appendix A.2).

Q3: What is the biosynthetic gene cluster composition of two species which are important in the perfume industry?
In this study I explore the diversity of biosynthetic gene clusters in *Evernia prunastri* and

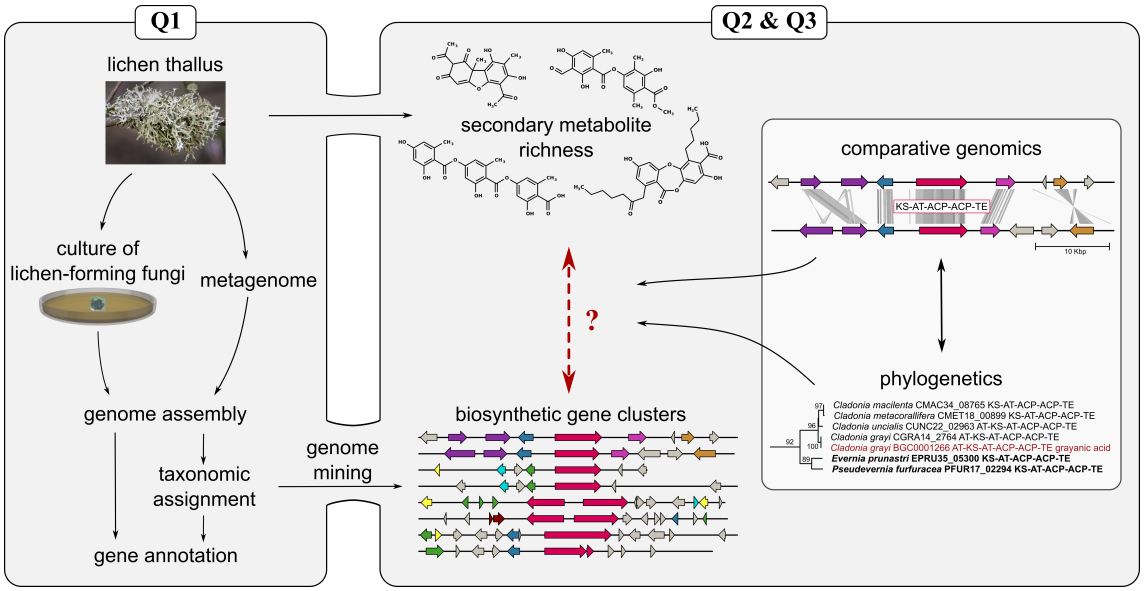


Figure 3.1. Thesis aim and structure. The overall thesis aim is the implementation of metagenomic approaches to better understand the biosynthetic gene content of lichen-forming fungi. This provides the basis for linking genes to secondary metabolites (highlighted in red). The research steps include validating a metagenomic approach to reconstruct genomes and gene content of lichen-forming fungi (**Q1**) and the exploration of the biosynthetic gene cluster content in two closely related sister-species (**Q2**) and in two metabolite-rich species (**Q3**). The genomes of lichen-forming fungi are the source for mining biosynthetic gene clusters. Comparative genomics and phylogenetics are used to bridge the missing link between genes and secondary metabolites.

Pseudevernia furfuracea and compare it to other lichenized fungi and non-lichenized fungi. I investigate the architecture and gene content of those clusters with high homology between *E. prunastri* and *P. furfuracea* and group polyketide synthases (PKSs) from *E. prunastri* and *P. furfuracea* phylogenetically in a phylogeny of PKSs with known functions. Results of this study are published in Calchera *et al.* (2019, see Appendix A.3).

4. Results and discussion

4.1. Metagenomic reconstruction of lichen-forming fungal genomes

Lichens' diverse and bioactive secondary metabolites are a promising source for new natural products leads especially in the light of growing resistances against common drugs (Boustie & Grube, 2005; Crawford, 2015). However, lichens are not amenable yet for biotechnological applications, mainly because of two reasons: i) the symbiotic nature of lichens make experimental strategies difficult or unfeasible, and ii) genetic information on the mycobiont, which is responsible for the biosynthesis of the lichen compounds of interest, is still scarce (Grube *et al.*, 2013; Grube & Wedin, 2016; Werth *et al.*, 2015). High-throughput sequencing of metagenomic samples became a promising approach for reconstructing genomes of symbionts which are difficult to cultivate or cannot be cultivated at all (Ghurye *et al.*, 2016). However, metagenomic approaches come with major analytical challenges mainly centered around two key issues: i) assembly strategy: finding an assembly strategy that fits the individual set-up, for example dealing with extreme coverage biases, and ii) taxonomic binning: applying a suitable assignment method to retrieve the genetic information of interest in high quality and without contamination of any other co-occurring organisms in the DNA mix (Greshake *et al.*, 2016; Tringe & Rubin, 2005). The applicability of such a metagenomic approach to assemble genomes of lichen-forming fungi from metagenomically sequenced samples has not been evaluated yet, but is crucial to advance natural product research on lichens.

In the first part of my thesis, I compared the genomes retrieved from a single shotgun sequencing of mixed DNA from natural lichen thalli with reference genomes assembled from DNA of pure fungal, aposymbiotic cultures. I chose the two species, *Evernia prunastri* and *Pseudevernia furfuracea*, which are particularly rich in secondary metabolites, and are two of the few lichen species to be harvested in large amounts (Joulain &

Tabacchi, 2009a,b). Furthermore, it was possible to obtain a mycobiont culture of both species with sufficient biomass for genome sequencing.

Reference genome assemblies and gene sets from culture-derived genomes

The reference genomes were assembled from sequencing read sets of paired-end and mate-pair libraries. Gene sets were predicted with the help of RNA-based evidence from metatranscriptome sequencing. Both mycobiont reference genomes were assembled to a similar total genome size and fungal genome completeness, the genome of *P. furfuracea* resulting into fewer and larger scaffolds with higher N50 and lower average coverage (Table 4.1). The more continuous genome of *P. furfuracea* is most likely the result of the better scaffolding, due to the larger mate-pair libraries, that we were able to obtain. A little more than half of the RNA-reads aligned to each of the genomes, and resulted in 10,922 genes predicted for *E. prunastri* and 8,842 genes predicted for *P. furfuracea*, both with a comparable fungal annotation completeness (Table 4.1). These culture-derived assemblies and gene sets represent the reference genomes and reference gene sets that were used for the comparison against the metagenome-derived assemblies.

Metagenomic assembly strategies

Genome assembly procedures follow no gold standard and require a certain extent of testing different approaches according to the specific project target, especially in organismal groups for which reference data is still scarce (Lantz *et al.*, 2018). I evaluated six different *de novo* assemblers and two taxonomic assignment methods for reconstructing the genomes of lichen-forming fungi from mixed DNA samples. For this I compared the quality and completeness of the genomes assembled from metagenome shotgun sequencing of natural lichen thalli with their respective fungal culture-derived genome. I included a variety of assembler types including general purpose as well as specialized metagenome assemblers, and assemblers utilizing overlap-layout-consensus or *de Bruijn* graphs (Nagarajan & Pop, 2013). The performance of assemblers was evaluated after taxonomic binning to find the method that is most suitable to reconstruct the fungal portion of the lichen. I assessed overall genome statistics, overlap to the reference genome, and fungal genome and gene set completeness. Overall, I found substantial differences depending on the assembler used, while the choice of taxonomic assignment method made no significant difference.

Table 4.1. Summary of genome and gene set comparisons of metagenome-assembled genomes compared to reference genomes of pure fungal culture strains.

	Evernia prunastri		*Pseudevernia furfuracea*	
	Culture-derived genome	Metagenome-derived genome[1]	Culture-derived genome	Metagenome-derived genome[1]
Number of scaffolds/contigs	277	1,775	46	3,558
Total size	40 Mb	39 Mb	38 Mb	43 Mb
N50	264 kb	53 kb	1,179 kb	36 kb
Average coverage	410x	110x	350x	135x
Number of genes	10,922	11,098	8,842	10,028
Orthologous genes[2]	9,915 (90 %)		7,763 (88 %)	
Fungal genome completeness[3]	95.9 %	95.4 %	94.7 %	93.7 %
Fungal gene set completeness[3]	92.1 %	91.7 %	91.8 %	91.3 %
Number of biosynthetic genes[4]				
Reducing type I PKS	20	20	12	9
Non-reducing type I PKS	8	8	5	6
Type III PKS	2	2	2	2
Hybrid PKS-NRPS	4	3	3	3
NRPS	4	4	4	5
Terpene synthases	12	14	5	6
Total number	50	51	31	31
Orthologous genes[2]	43 (86 %)		24 (77 %)	

[1] The metagenome-assembled genomes were generated with the assembler SPAdes and taxonomically assigned with MEGAN.
[2] Orthologous gene pairs between the reference and the metagenome-assembled genomes were identified by Reciprocal Best Blast Hits (RBH).
[3] Fungal genome and gene set completeness was estimated against a lineage-specific set of 1,315 Ascomycota single-copy orthologs with BUSCO and the percentage of complete orthologs are given in the table.
[4] Secondary metabolite biosynthetic genes were identified with antiSMASH.

The evaluation of the fungal assembly quality was based on several statistics such as number of contigs, largest contig, total length, Nx values (e.g. N50), number of misassemblies and the fraction of reference genome and genes that were covered. Genomes assembled by SPAdes (Bankevich *et al.*, 2012) showed the best overall assembly statistics and the highest overlap to the reference genomes (86–90 % for *E. prunastri* and 80–87 % for *P. furfuracea*). The dedicated metagenome assembler metaSPAdes (Nurk *et al.*, 2017) performed also well, but only second best. Some of the tested assemblers including dedicated metagenome assemblers such as omega (Haider *et al.*, 2014) and metaVelvet (Namiki *et al.*, 2012) generated only poor assemblies and had less than 75 % overlap to the reference. The reason for the differences in assembly quality may lie in the presence of highly non-uniform read coverages among different genomes present in the sample (Nurk *et al.*, 2017). The lichen thallus represents a mix of species from

different kingdoms including fungi, algae and bacteria. As such, metagenomic read sets obtained from lichen thalli were shown to cause particular challenges for assemblers, particularly because of uneven coverages (Greshake *et al.*, 2016). Greshake Tzovaras *et al.* (2019) reported a coverage variation in the holo-genome of the lichen *Umbilicaria pustulata* extending over three orders of magnitude among various genomes. Some assemblers such as omega and metaVelvet make use of coverage information to differentiate between species. In the case of lichen metagenomes, this sensitivity to uneven coverage may effect the fungal genome assembly from short-read sequences negatively (Pop, 2009). The best performing assemblers were to a certain extent developed to deal with non-uniform sequencing data.

Taxonomic assignment

I tested two different approaches for taxonomic classification to retrieve the fungal portion of the metagenome-derived contigs. The first approach is based on sequence similarity through a translated nucleotide BLAST search against the entire NCBI Genbank non-redundant protein database. The results are parsed to MEGAN and each sequence is assigned to the lowest common ancestor (LCA) taxon in the NCBI taxonomy based on a weighted algorithm for increased sensitivity (Huson *et al.*, 2016). All contigs assigned to Ascomycota were considered to represent the fungal mycobiont since not enough closely related reference genomes are available yet (Spribille *et al.*, 2016). The second approach uses taxonomic profiling together with multivariate statistics of tetranucleotide frequencies and differential coverage binning through the program MetaWatt (Strous *et al.*, 2012). All bins with at least 50 % of the fragments classified as Ascomycota and without other taxa in the taxonomic profile were merged to obtain the sequences corresponding to the mycobiont. The choice of taxonomic assignment method made no significant difference, but the BLAST/MEGAN approach performed slightly better when comparing the overall genome statistics and the overlap to the reference. However, the approaches differed in computational needs. The BLAST sequence similarity approach for MEGAN required a larger database and more computational resources than MetaWatt. This latter method can use a smaller database and efficiently deploys multivariate statistics of tetranucleotide frequencies and differential coverage binning for the assignment.

Additionally, I estimated fungal genome and gene set completeness based on evolutionarily-informed expectations of gene content and a lineage-specific set of Ascomycota

single-copy orthologs (Simão *et al.*, 2015). By comparing the completeness in the assembled contigs before and after taxonomic assignment to Ascomycota, I was able to track if potential fungal contigs were lost during the assignment process. In the case of the better assembled scenario (i.e., fewer and larger scaffolds) the fungal genome completeness of the unassigned contigs were similar to the completeness in the contigs assigned taxonomically to Ascomycota. However, for the less well assembled contigs, the fungal genome completeness was lower suggesting a loss of fungal contigs during the assignment process. This may stem from the fact that shorter sequences are more difficult to assign taxonomically (Greshake Tzovaras *et al.*, 2019; Sczyrba *et al.*, 2017; Vollmers *et al.*, 2017; Zhang, 2018). In general, fungal genome completeness was comparable between assignment methods, but MEGAN-derived assemblies had slightly higher mycobiont completeness and assigned a higher number of contigs to Ascomycota. SPAdes showed the highest fungal genome completeness in MEGAN-assigned assemblies and resulted in 95.5 % for *E. prunastri* and 94.0 % for *P. furfuracea* (Table 4.1).

Combined evaluation

Overall, the evaluation of assembly statistics, overlap to the reference genome, and the assessment of fungal genome completeness confirmed SPAdes and MEGAN as the best suitable strategy to retrieve the genome of the mycobiont from metagenomic samples in both investigated species. The metagenome-derived assemblies show a similar total genome size and fungal genome completeness with a lower average coverage compared to the reference genomes. However, the metagenomic assemblies were more fragmented as indicated by the lower N50 and the higher number of scaffolds in comparison to the reference assemblies (Table 4.1). This can however be expected from the single paired-end shotgun sequencing design for the metagenomic samples in comparison to the reference assemblies that were sequenced with different paired-end and mate-pair sequencing libraries at a deeper coverage. Our results, however, indicate that a reference genome is not necessarily required to choose the best metagenomic assembly and taxonomic assignment strategy. Evaluating the assemblies based on assembler statistics and fungal genome completeness without reference statistics such as overlap to the reference genome would have led to the same conclusion as a reference-based evaluation. The few other studies available to date that have assembled mycobiont genomes from metagenomic DNA of lichen thalli also generated fragmented assemblies with more than thousand scaffolds (Allen *et al.*, 2018; Dal Grande *et al.*, 2018a, 2017; McDonald

et al., 2013b). The well-assembled lichen holo-genome of *Umbilicaria pustulata* from metagenomic shotgun data was based instead on a combination of Illumina and long PacBio reads, thus highlighting the potential of long-read data to overcome the above mentioned challenges in lichen metagenomics (Greshake Tzovaras, 2017; Greshake Tzovaras *et al.*, 2019).

Taxonomic composition of metagenomic reads sets

To assess the effect of non-mycobiont DNA on the ability to reconstruct the mycobiont genome from metagenomic samples, I estimated the proportion of the targeted mycobiont in the metagenomic DNA. For this I investigated the taxonomic composition of all metagenomic reads from *E. prunastri* and *P. furfuracea*. I taxonomically classified all trimmed and corrected reads using exact alignments of overlapping *k*-mers (Lindgreen *et al.*, 2016; Wood & Salzberg, 2014). The advantage here is the incorporation of the reference genomes of both investigated species in a custom database. This enables a precise estimation of the read amounts of the lichen-forming fungi and potential other fungal species present in the metagenome. Besides the reference genomes of *E. prunastri* and *P. furfuracea*, the database included the entire fungal RefSeq, some genomes of lichen-forming fungi from NCBI and the basidiomycete *Cystobasidium pallidum* (Spribille *et al.*, 2016). The mycobiont of *E. prunastri* represented 73.4 % of the metagenomic reads and the mycobiont of *P. furfuracea* 72.2 %. Cystobasidiomycete yeasts were not identified in the metagenomic reads of *E. prunastri* and *P. furfuracea*. This is in line with the recent finding that these yeasts may not be ubiquitous in lichens (Lendemer *et al.*, 2019) in contrast to what has been suggested by Spribille *et al.* (2016). Further sources of contamination were assessed by taxonomically classifying all metagenomic reads that were not classified as reference mycobiont with a BLAST/MEGAN assignment based on the entire NCBI protein database. Besides the mycobiont, 10.7 % Bacteria, 0.6 % Viridiplantae, and 0.7 % Fungi were identified in *E. prunastri* and 1.9 % Bacteria, 2.1 % Viridiplantae and 1 % Fungi in *P. furfuracea*. The remaining proportion of reads was of unknown origin and was substantially higher in *P. furfuracea*.

Comparison of gene space

I compared the gene sets of the metagenome-derived genomes of both taxonomic assignment methods with the gene sets of the reference genomes. Therefore I first

performed a *de novo* gene prediction using RNA-based evidence and assessed the fungal completeness of the resulting gene sets. Again more than half of the RNA-reads aligned to the metagenome-derived assemblies and provided hints for gene finding. I predicted 11,098 and 10,713 genes in the MEGAN- and MetaWatt-assigned assemblies of *E. prunastri* and 10,028 and 8,962 genes in the MEGAN- and MetaWatt-assigned assemblies of *P. furfuracea*. All four metagenome-derived fungal gene sets had a fungal genome completeness between 89.2 % and 91.7 % comparable to the reference gene sets (Table 4.1).

Next, as a simple and fast method to compare the overlap of gene sets from different assemblies of the same species, I identified Reciprocal Best Blast Hits (RBH) to find orthologous pairs between the gene sets (Moreno-Hagelsieb & Latimer, 2008; Ward & Moreno-Hagelsieb, 2014). The identified orthologous pairs between the assemblies of both taxonomic assignment methods and the reference gene sets overlapped to a great extent in both species. The core genes present in all three sets – MEGAN-assigned, MetaWatt-assigned and reference – reached 87 % in *E. prunastri* and 83 % in *P. furfuracea*. The metagenomic gene set from MEGAN covered in both species more of the reference than the gene sets of MetaWatt. Overall, 88 - 90 % of the respective reference gene space was covered with the MEGAN-derived metagenomic gene sets (Table 4.1).

After showing that most of the fungal gene space overlaps between reference and metagenome-based assemblies, I tested if the metagenome-based assemblies are also reliable for studying the gene families of secondary metabolism. Owing to the richness and uniqueness of secondary metabolites found in lichen-forming fungi, these gene families are especially interesting targets for focused genome mining studies (Abdel-Hameed *et al.*, 2016b; Huneck, 1999). For genome-wide identification, annotation and analysis of secondary metabolite gene clusters the widely-used automated genome mining pipeline antiSMASH was used (Blin *et al.*, 2017). I identified a high number of PKSs, NRPSs, and terpene synthases in both species and found that many biosynthetic genes found in culture overlap with the genes found in the MEGAN- and MetaWatt-assigned assemblies (Table 4.1). The culture-based reference genome of *E. prunastri* contained 50 secondary metabolite gene clusters of which 41 were present in both MEGAN- and MetaWatt-assigned metagenome assemblies. In *P. furfuracea* 22 of 31 biosynthetic gene clusters found in the reference assembly were also found in the metagenome-derived assemblies. MEGAN-assigned assemblies contained more biosynthetic genes than the MetaWatt-assigned assemblies in both species (51 in *E. prunastri*

and 31 in *P. furfuracea*). Metagenome-derived assemblies recovered most of the diverse secondary metabolite gene families found in the reference genomes. This highlights the potential of metagenome skimming for natural product discovery in lichens (Adnani *et al.*, 2017). Furthermore, these results indicate that the metagenomic skimming approach is extendable for whole genome mining of other target gene families, e.g. symbiosis-related proteins, mating-type genes, or secreted effector proteins (Wang *et al.*, 2014b).

Presence of potential fungal contaminants

A few secondary metabolite genes were only found in the metagenome and not in the reference. These genes may originate from the annotation of more fragmented metagenome-assembled genomes or may be missing in the culture-assembled reference genome. Since these gene clusters are of fungal origin another possibility includes contaminations from other non-lichen-forming ascomycetes (i.e., lichenicolous fungi). However, only a small fraction of additional fungal reads was found besides the primary mycobiont in the assessment of taxonomic composition of the metagenomic reads. This fraction of non-mycobiont ascomycete reads was assembled into contigs and searched for the presence of biosynthetic genes. One non-mycobiont NRPS in the metagenome of *E. prunastri* and one hybrid PKS-NRPS not belonging to the mycobiont of *P. furfuracea* were found. But almost three-fourths of the reads belonged to the primary mycobiont while less than one percent was classified to other non-mycobiont fungi. Therefore only a minor proportion of fungal contaminations may be present in the metagenomes and likely does not infer with our ability to reconstruct lichen-forming fungi from metagenomic samples. Essential here is the sensitivity and precision of the taxonomic binning of sequences and the availability and quality of reference databases (Nasko *et al.*, 2018). Without good reference databases it is not yet possible to exclude contamination from metagenomic assemblies entirely.

Summary – Metagenomic reconstructions

Metagenome sequencing is a valuable approach to gain insights into the genomes of lichen-forming fungi circumventing culturing. No study has evaluated the accuracy and completeness of assemblies based on metagenomic sequences in comparison with assemblies based on cultured strains of lichenized fungi. Therefore I first generated

and annotated the two reference genomes of *E. prunastri* and *P. furfuracea* from fungal cultures, which contributes to the few lichen genomes available to date. I identified the best suitable metagenomic assembly and taxonomic assignment strategies for both species, assessed the taxonomic composition of the metagenomic reads and compared the gene space of the metagenome-derived assemblies with the reference. Specifically, I showed that:

- Metagenome-derived fungal assemblies are comparable to culture-derived reference genomes in terms of total genome size and fungal genome completeness.
- The choice of assembler is essential, while the tested taxonomic assignment methods had no large effect on the quality of the assembled genomes.
- Less fragmented assemblies recovered a higher fungal proportion during the taxonomic assignment.
- A suitable metagenomic assembling strategy can be inferred without reference genome.
- Three-fourths of the metagenomic sequencing data belonged to the primary mycobiont, the remaining portion can vary in taxonomic content between species.
- Non-mycobiont DNA does not effect the reconstruction of the lichen-forming fungi strongly.
- Fungal gene space and fungal gene completeness are comparable in culture-derived and metagenome-derived genomes.
- Metagenome-derived assemblies reliably recover gene families of secondary metabolism, but potentially also other groups of gene families.

4.2. Biosynthetic gene clusters of lichen-forming fungi

The aim of this thesis is to better understand the biosynthetic gene content of lichen-forming fungi to eventually link genes to secondary metabolites, and provide a basis for biotechnological approaches to exploit lichen secondary metabolism. Furthermore, broadening our understanding of the biosynthetic genes in lichenized fungi may help in gaining insights in the evolution of these important gene families. The high number of cryptic or silent gene cluster compared to the known number of metabolites hints at the potential production of many unknown metabolites yet to be detected (Bertrand *et al.*, 2018a). These undiscovered metabolites may play essential roles in maintaining the lichen symbiosis or in the communication between symbionts, or may be only relevant in certain life stages, for example while finding compatible symbiont partners

in the environment and during thallus formation (Calcott *et al.*, 2018). However, these lichen-derived metabolites are hitherto inaccessible and their roles are therefore not understood.

4.2.1. Two closely related sister-species

After I showed that metagenomic sequencing can be used to obtain the genomes of lichen-forming fungi, I was able to assess and compare the biosynthetic gene content of two closely related species, one assembled from metagenomic DNA (Greshake Tzovaras, 2017) and one assembled from an available mycobiont culture (Dal Grande *et al.*, 2018a). We generated a *de novo* genome assembly of *Umbilicaria hispanica* and compared it to the closely related, but ecologically distant *Umbilicaria pustulata*. Until recently both species belonged to the genus *Lasallia*, but recently *Lasallia* was reduced to synonymy with *Umbilicaria*, the only genus in the family of Umbilicariaceae (Davydov *et al.*, 2017). Both are European species, but while *U. pustulata* has a wide distribution from the Mediterranean to boreal-montane habitats, *U. hispanica* is endemic to sky island habitats in the Mediterranean region (Hestmark, 1992; Sancho & Crespo, 1989; Vivas *et al.*, 2017). The species differ in their water acquisition strategies. *U hispanica* takes up moisture directly from fog and low-lying clouds whereas *U. pustulata* relies on surface water run-offs (Vivas *et al.*, 2017). *Umbilicaria hispanica* predominantly reproduces sexually, while *U. pustulata* reproduces mainly vegetatively by isidia (Sancho & Crespo, 1989). Despite different ecological requirements, the species show similar chemical profiles, with gyrophoric acid as the major compound, lecanoric acid, hiascic acid and skyrin as minor compounds, and traces of umbilicaric acid (Posner *et al.*, 1991). I analysed the biosynthetic gene cluster content of these sister-species to investigate if their taxonomic proximity is reflected in their chemical profile.

Genome assemblies

DNA from pure mycobiont culture of *U. hispanica* was sequenced on a single Illumina short insert paired-end library (300 bp) and two long insert mate-pair libraries (3 kb and 8 kb). After adapter and quality trimming, sequences were assembled and scaffolded, and assigned taxonomically to Ascomycota with multivariate statistics of tetranucleotide frequencies and differential coverage binning to filter out potential contaminants. The resulting 1,619 scaffolds had a total sequence length of 41 Mb, a N50 of 145 kb and

an average coverage of 160x. An estimated completeness of 96.4 % indicated that the genome space was well covered (BUSCO v3.0.2, Simão *et al.*, 2015). We predicted a total of 8,488 genes with the help of training data from RNA sequences of *U. pustulata*, and protein sequences of *Cladonia grayi* and *Xanthoria parietina*. The predicted gene set had a rather moderate fungal gene space completeness of 82.9 % and was annotated with gene ontology (GO) terms and protein names.

The genome assembly of *U. pustulata* was obtained from a hybrid assembly of Illumina and PacBio reads done by Greshake Tzovaras *et al.* (2019). The *U. pustulata* genome assembly consisted of 43 scaffolds of a total length of 33 Mb, a N50 of 1,808 kb and included 9,825 genes. The genome assembly had an estimated genome completeness of 93.6 % and gene set completeness of 91.9 % (BUSCO v3.0.2).

Synteny and orthology analysis

Analogously to the comparison of gene space in section 4.1 on page 23 (Meiser *et al.*, 2017), we identified Reciprocal Best Blast Hits (RBH) for finding orthologs between the gene sets of the two closely related species. Based on this, we found 6,382 orthologous gene pairs between *U. hispanica* and *U. pustulata*. This means that 75 % of the *U. hispanica* genes are orthologs to *U. pustulata*. Furthermore, the genome assemblies were compared to find syntenic regions with conserved gene order. Scaffolds longer than 50 kb of both assemblies were aligned and ordered by computed syntenic order (Kurtz *et al.*, 2004; Soderlund *et al.*, 2011). This led to 68 % syntenic coverage in *U. hispanica* and 71 % in *U. pustulata*. In *U. hispanica* the aligned scaffolds represented 31.1 Mb of the genome in 211 scaffolds, and in *U. pustulata* only 31 scaffolds totalling 32.7 Mb. More than 80 % of genes were in syntenic blocks in both species. Despite the less contiguous genome of *U. hispanica*, a high degree of synteny conservation between the two species with only few rearrangements could be shown.

Biosynthetic gene composition

Secondary metabolite biosynthetic gene clusters in *U. hispanica* and *U. pustulata* were identified and annotated with the fungal version of antiSMASH (Blin *et al.*, 2017). Predicted gene families included PKS, NRPS, hybrid PKS-NRPS and terpene synthases and were compared in terms of total gene number and RBHs. In total, 18 biosynthetic gene clusters with complete core genes were predicted in *U. hispanica* and 21 in

U. pustulata. Two clusters had a putative orthologous core gene, but no homology between the genes in the cluster. Eleven clusters showed a high similarity not only in the core gene but also in tailoring genes (compare Table S3 on page 87 in Appendix A.2). These orthologous gene clusters are of particular interest as they may encode for secondary metabolites common to both species, such as the major compound gyrophoric acid. On the other hand, clusters that differ between the *Umbilicaria* species might be relevant to gain insights in the biosynthesis of metabolites underlying ecological and reproductive differences between the species. In general, our results highlight the potential of both species to produce more metabolites than expected from their rather poor chemical profile compared to other lichenized fungi. The genomics-driven discovery of biosynthetic gene clusters together with chemical profiling and comparative genomics of gene clusters can guide the first steps towards linking natural products with the gene clusters encoding them (Chooi & Solomon, 2014).

4.2.2. Two metabolite-rich species

In the last part of my thesis, I compared the biosynthetic gene inventory of the two lichen-forming fungi *Evernia prunastri* and *Pseudevernia furfuracea* which are rich in metabolites and have substantial commercial value in the perfume industry. I used the culture-derived genome assemblies from the first part of my thesis for an in-depth characterization of the architecture and gene content of metabolic gene cluster with high homology between the two species. The diversity of secondary metabolite gene clusters of the two species was set into a larger context by comparing it against the genomes of other lichenized and non-lichenized fungi. Furthermore, I used a phylogenetic approach to gain functional hints for the lichen-derived biosynthetic genes through relatedness to a set of functionally characterized fungal secondary metabolite gene clusters.

Biosynthetic gene richness in genomes of lichenized and non-lichenized fungi

Chemical diversity and biosynthetic gene richness of *E. prunastri* and *P. furfuracea* were surveyed and compared to all available genome assemblies of other lichen-forming fungi. The genomes span a range of phylogenetic groups and synthesized chemical profiles. All assemblies are from fungal culture except for the metagenome-derived assembly of *U. pustulata*. Here I used a previous version of the *U. pustulata* genome assembly that

was already available in a public database and based only on Illumina data in contrast to the later updated assembly that became only very recently publicly available and includes PacBio data (Greshake Tzovaras *et al.*, 2019). The assembly status differed considerably among genomes, but most gene spaces were well covered according to genome completenesses estimates of 90–96 % with only two exceptions (Table S1 on page 114 in Appendix A.3). Seven genomes required a *de novo* gene prediction as it greatly improves biosynthetic gene cluster detection. No transcriptome data was available for these genomes resulting in less complete (72–80 %) and more fragmented gene sets in comparison to gene sets predicted with RNA evidence (82–97 %). However, these gene sets still represent better proxies than the annotation of the secondary metabolite detection software (Blin *et al.*, 2017).

A high number of secondary metabolite gene clusters was identified with antiSMASH in all fifteen genomes of lichen-forming fungi – 47 clusters per genome on average. Highest richness was predicted for *E. prunastri* with 80 biosynthetic gene clusters; 51 gene clusters were predicted for *P. furfuracea*. The most abundant gene families in all genomes were R-PKS, although most polyketides found in lichens are reported to be fully oxidized (Stocker-Wörgötter, 2015). Other high abundant gene families are terpene synthases and NR-PKS. The diversity of R-PKS found in lichen genomes suggests the presence of a high, so far untapped, biosynthetic potential in lichenized fungi.

Furthermore, I compared the number of lichen secondary metabolic gene clusters with those from 57 genomes representing all major fungal classes (Rokas *et al.*, 2018). The genomes of lichenized fungi in the Lecanoromycetes harboured a particularly high biosynthetic gene richness compared to genomes of non-lichenized fungi. In total, the average number of 11 PKS clusters was exceeded by almost all lichen genomes which on average contained 24 PKS gene clusters. The highest number of PKS clusters in the data set was found for the lichen *E. prunastri*. It has to be noted, however, that my analysis does not represent an exhaustive comparison of all representative fungal genomes as other fungal species rich in secondary metabolite gene clusters have been reported (Wang *et al.*, 2015). Nevertheless, my analysis highlights the high biosynthetic potential that is hidden in the genomes of many lichenized fungi.

In-depth gene cluster comparison

Evernia prunastri and *P. furfuracea* revealed a high number of biosynthetic gene clusters and both were shown to harbour many chemical constituents (Joulain & Tabacchi, 2009a,b). Both species synthesize atranorin and chloroatranorin, which have potential as drugs (Boustie & Grube, 2005; Zhou *et al.*, 2017). Therefore a comparative cluster mapping may detect biosynthetic enzymes that synthesize characteristic lichen substance classes. For the in-depth comparison of biosynthetic gene cluster between *E. prunastri* and *P. furfuracea*, I applied a two-step approach. RBH were used as proxy for detecting orthologous pairs between all cluster genes of both species. In 887 cluster genes in *E. prunastri* and 548 cluster genes in *P. furfuracea*, I identified 126 orthologous pairs. All genes belonging to a biosynthetic gene cluster were functionally annotated with GO terms and protein names to include all potential accessory genes involved in metabolite synthesis. Next, I analysed the homology and arrangement of entire gene clusters through comparative mapping. All PKS, NRPS and hybrid PKS-NRPS gene clusters with a RBH core gene were selected and their synteny plotted based on a translated nucleotide query BLAST against a translated nucleotide database.

Twelve gene clusters had a RBH core gene and nine of these clusters showed a high cluster homology between the two species. Several accessory cluster genes of these highly homologous gene clusters were also RBHs and carried similar functional annotations for example cytochrome P450, monooxygenases, serine/threonine kinases, and regulatory genes. The homologous regions of the gene clusters were mainly confined within genes. A few exceptions with homology between non-coding regions may indicate genes missed during in gene prediction. In total, three out of four NR-PKS gene clusters, two of four R-PKS gene clusters and all NRPS and hybrid PKS-NRPS gene clusters showed a cluster homology between *E. prunastri* and *P. furfuracea*.

A manual protein BLAST indicated the presence of starter acyl carrier protein transacylase (SAT) and product template (PT) domains in the domain architecture of NR-PKS genes. Both domains are characteristic of NR-PKS but were not annotated by antiSMASH (Cacho *et al.*, 2015; Crawford *et al.*, 2006, 2009, 2008). A few core NR-PKS genes were split over two or three genes in only one or both species. This may result from artefacts in gene prediction or represent gene clusters that are functionally interrupted, especially for those clusters with a high homology. I also detected a NRPS and a hybrid PKS-NRPS encoded closely together, which might suggest that the two natural products function together or that both enzymes are required for the synthesis

of one natural product. Some clusters have been predicted at scaffold ends and limit the comparison. Long read sequencing could greatly improve the continuity of genome assemblies and overcome these issues. However, it remains a challenge in slow growing symbiotic organisms to obtain sufficient culture- or metagenome- DNA for long read sequencing. Further improvements in sequencing technology and protocol procedures optimized for low input DNA will aid in making symbiotic systems more accessible for genomics-driven natural product research.

Putative regulators of biosynthetic gene clusters

The transcriptional control of biosynthetic gene cluster is complex and can exist on several levels, from global to pathway-specific regulation (i.e. intracluster) (Brakhage, 2013). Often specific acting gene regulation is found within the clusters that these transcription factors regulate. I identified putative biosynthetic genes regulators from the functional annotation of cluster genes in *E. prunastri* and *P. furfuracea*. Most putative intracluster regulators were found in *E. prunastri* (44 of 60). Two sets of orthologous regulator genes were identified between the species based on RBHs in a NR-PKS gene cluster and in a NRPS/hybrid PKS-NRPS gene cluster. The presence of intracluster regulators suggest that their co-expression may be required for successful heterologous expression attempts.

Phylogenetic analysis with characterized fungal polyketides

Combining genome mining with phylogenetic analysis may provide useful hints for the functional characterization of characteristic lichen substance classes (Blin *et al.*, 2019; Jenke-Kodama *et al.*, 2005; Nielsen *et al.*, 2017). For assessing the phylogenetic relatedness of lichen biosynthetic genes to fully functionally characterized secondary metabolite genes, I compiled a custom database with 131 entries from a repository of experimentally characterized fungal metabolites (Medema *et al.*, 2015). I compared a total of 413 ketoacyl synthase domain (KS) protein sequences of all PKS and hybrid PKS-NRPS genes identified in the fifteen annotated genomes of lichen-forming fungi containing at least three PKS domains and long enough KS sequences. The KS domain is an essential part of PKS genes and has been extensively targeted by previous phylogenetic studies due to its relatively high level of sequence conservation, in particular at the amino acid level (Kroken *et al.*, 2003). Furthermore, the KS domain has been

shown to be highly predictive in inferring domain architecture and pathway association (Jenke-Kodama *et al.*, 2005; Ziemert & Jensen, 2012). Database and lichen-derived KS sequences were aligned and placed in a maximum likelihood phylogenetic tree using animal fatty acid sequences as outgroup (Darriba *et al.*, 2011; Katoh, 2002; Katoh & Standley, 2013).

The phylogenetic tree comprised two major clades – one clade with R-PKSs and one clade with NR-PKSs. Biosynthetic gene clusters of *E. prunastri* and *P. furfuracea* were spread across the phylogeny (Fig. S1 on page 126 in Appendix A.3). I further analysed in detail four NR-PKS clades, as they are considered the building units of lichen-specific depsides and depsidones (Calcott *et al.*, 2018; Elix & Stocker-Wörgötter, 2008; Stocker-Wörgötter, 2008). A PKS gene from *E. prunastri* grouped within a supported subclade of C-methyltransferase domains including the methylphloroacetophenone synthase (*MPAS*) gene of *Cladonia uncialis* that is putatively linked to usnic acid biosynthesis. Interestingly, usnic acid is a secondary metabolite also reported for *E. prunastri* (Abdel-Hameed *et al.*, 2016b). In another clade two NR-PKS from the highly homologous gene cluster of *E. prunastri* and *P. furfuracea* fell phylogenetically close to the PKS gene for grayanic acid production in *Cladonia grayi* (Armaleo *et al.*, 2011). Interestingly, all four *Cladonia* species included in this study have a PKS in this clade, further corroborating the evidence that the genes are responsible for the biosynthesis of grayanic acid or in the cases of *E. prunastri* and *P. furfuracea* very similar compounds. Two closely related NR-PKS of *E. prunastri* and *P. furfuracea* fell within a group of tandem acyl-carrier protein (ACP) domains comprising hydroxy naphthalenes genes (Fulton *et al.*, 1999; Zhang *et al.*, 2003). This substance group was reported in extracts of both species and is used in the perfume industry (Stocker-Wörgötter, 2008). The last example of an interesting NR-PKS includes two genes from *E. prunastri* and *P. furfuracea* grouping with the gene *orsA*, that was experimentally characterized to be involved in the orsellinic acid biosynthesis in *Aspergillus nidulans* (Sanchez *et al.*, 2010). Orsellinate type monoaromatic units are the basic building blocks of the most abundant classes of lichen metabolites (Calcott *et al.*, 2018). Combining comparative mapping of entire clusters with a phylogeny based solely on the conserved KS domain revealed that cluster similarity corresponds to KS topology, thus confirming the high predictive power of KS domains (Jenke-Kodama *et al.*, 2005; Kroken *et al.*, 2003; Ziemert & Jensen, 2012). These examples I reported constitute promising candidates for linking secondary metabolites and metabolic gene clusters.

Summary – Biosynthetic gene clusters

Whole genome mining of biosynthetic gene clusters has revealed a large number of uncharacterised secondary metabolite gene clusters in the genomes of fifteen species of lichen-forming fungi. These numbers exceed by far the number of produced metabolites reported from these species and highlight a hidden biosynthetic potential. Furthermore, I provided an in-depth functional characterization of biosynthetic gene clusters in two promising systems: i.e. two ecological distinct *Umbilicaria* sister species and two metabolite-rich species that are important in the perfume industry. The structural and phylogenetic analysis of biosynthetic gene clusters provide new insights into lichen biosynthesis pathways and help progressing our understanding of the link between gene and compound. I combined genomics, comparative mapping and phylogenetics to show the following:

- The genomes of lichen-forming fungi show a high richness of secondary metabolic gene clusters compared to other fungal classes, especially in PKS gene clusters.
- The diversity of biosynthetic gene clusters in all fifteen genomes of lichen-forming fungi exceeds by far the number of reported metabolites.
- The species *Evernia prunastri* and *Pseudevernia furfuracea* – important in the perfume industry – show both a particularly high number of biosynthetic genes (80 gene clusters in *E. prunastri* and 51 in *P. furfuracea*).
- The ecologically distinct sister-species *Umbilicaria hispanica* and *U. pustulata* showed a high homology in eleven biosynthetic gene clusters in accordance with the high synteny in gene content and order that I found in both genomes.
- Phylogenetic analysis revealed promising candidate clusters grouping lichen metabolite genes with functionally characterized fungal polyketides.
- Whole gene cluster similarity corresponds to the topology of a phylogeny based on the conserved KS domain alone, this highlights its value for predicting putative domain architecture and pathway association.
- The identification of putative intracluster regulators will aid heterologous expression.
- The comparison of several genomes and gene sets of lichen-forming fungi demonstrate that assembly, together with RNA-based gene prediction, are crucial factors in determining accurate gene cluster assessments.

5. Summary and outlook

With my thesis I broadened our understanding of the biosynthetic gene content of lichen-forming fungi to open up research avenues into linking genes to secondary metabolites and unlock these organisms for biotechnological applications. In the first part, I evaluated the quality of metagenomically assembled genomes in lichen-forming fungi to validate this strategy for accessing the biosynthetic potential in symbiotic systems with challenges in cultivation. In the second part, I investigated specifically the biosynthetic gene content in two promising set-ups: first in the ecological distinct species *Umbilicaria hispanica* and *U. pustulata*, and last in the particularly metabolite-rich and economically relevant lichens *Evernia prunastri* and *Pseudevernia furfuracea*.

Q1: How well can we reconstruct the genomes and gene content of lichen-forming fungi from metagenomic samples?
Through a direct comparison of genome assemblies and gene sets obtained from natural, metagenomic lichen thalli with those from axenically cultivated mycobionts I could show that metagenome skimming is a viable approach to reconstruct the fungal genomes from uncultured lichen thalli and cover most of the gene space. I show that the choice of assembler is essential, while the methods of taxonomic assignment made no large difference. Furthermore, non-mycobiont sequences did not affect the ability to reconstruct the genomes of lichenized fungi from metagenomic samples. These findings indicate that metagenome-derived genome assemblies might work well for other mutualistic or parasitic/pathogenic system in the Eukaryotes.

Q2: What is the biosynthetic content of two sister-species with different ecological requirements?
I identified and compared the biosynthetic gene content of *U. hispanica* assembled from a mycobiont culture with the gene content of *U. pustulata* assembled from metagenomic DNA. Both species synthesize only the same few secondary metabolites despite their ecological, reproductive and genetic differences. I detected numerous gene clusters in both species suggesting a far greater potential to produce specialized secondary metabolites than expected from their chemical profile. Some biosynthetic gene clusters showed a high homology between both species in accordance with the high synteny in

gene content and order in both genomes. These clusters represent ideal candidates for secondary metabolites synthesized in both species, while the remaining clusters may encode for metabolites relevant for the different ecological requirements of both species. *U. hispanica* occurs in frost prone, high elevation environments. Therefore it would be interesting to investigate, which biosynthetic genes, and other genes might be linked to frost tolerance.

Q3: What is the biosynthetic gene cluster composition of two species which are important in the perfume industry?
Whole-genome mining revealed a remarkable biosynthetic richness in both species in comparison to other lichenized and non-lichenized fungi exceeding by far the number of reported metabolites. I applied comparative mapping for an in-depth characterization of architecture and gene content of clusters with high homology between both species. Furthermore, I identified promising candidate clusters from phylogenetic relatedness to functionally characterized genes. Both aid together with the identification of putative cluster regulators in providing new insights in lichen biosynthetic pathways and facilitate the development of biotechnological systems to tap the natural product richness discovered in lichens. It would be particularly interesting to identify the clusters associated with usnic acid and atranorin biosynthesis, two compounds with antibiotic and anti-inflammatory activities.

In conclusion, I showed that the metagenomic reconstruction of genome sequences from mixed-species samples is feasible to circumvent the time-consuming step of cultivation and could be applied to unculturable organisms or other complex communities. Whole-genome mining of gene families in combination with comparative mapping and phylogenetics provide an overview of the metabolic potential of a species. Furthermore, the new reference genomes of *E. prunastri*, *P. furfuracea* and *U. hispanica* constitute a valuable resource for facilitating comprehensive genome mining, comparative (phylo)genomics, the development of high-resolution genetic molecular markers, or populations genetics. Specifically will my work aid in gaining new insight into lichen biosynthetic pathways and serve as valuable resource for developing heterologous expression systems to tap the remarkable secondary metabolite richness of lichens in biotechnological applications.

References

Abdel-Hameed, M., Bertrand, R. L., Piercey-Normore, M. D., & Sorensen, J. L. (2016a). Identification of 6-hydroxymellein synthase and accessory genes in the lichen *Cladonia uncialis*. *Journal of Natural Products*, 79(6), 1645–1650, doi: 10.1021/acs.jnatprod.6b00257.

Abdel-Hameed, M., Bertrand, R. L., Piercey-Normore, M. D., & Sorensen, J. L. (2016b). Putative identification of the usnic acid biosynthetic gene cluster by *de novo* whole-genome sequencing of a lichen-forming fungus. *Fungal Biology*, 120(3), 306–316, doi: 10.1016/j.funbio.2015.10.009.

Adnani, N., Rajski, S. R., & Bugni, T. S. (2017). Symbiosis-inspired approaches to antibiotic discovery. *Natural Product Reports*, 34(7), 784–814, doi: 10.1039/C7NP00009J.

Allen, J. L., McKenzie, S. K., Sleith, R. S., & Alter, S. E. (2018). First genome-wide analysis of the endangered, endemic lichen *Cetradonia linearis* reveals isolation by distance and strong population structure. *American Journal of Botany*, 105(9), 1556–1567, doi: 10.1002/ajb2.1150.

Armaleo, D., Müller, O., Lutzoni, F., Andrésson, Ó. S., Blanc, G., Bode, H. B., Collart, F. R., Dal Grande, F., Dietrich, F., Grigoriev, I. V., Joneson, S., Kuo, A., Larsen, P. E., Logsdon, J. M., Lopez, D., Martin, F., May, S. P., McDonald, T. R., Merchant, S. S., Miao, V., Morin, E., Oono, R., Pellegrini, M., Rubinstein, N., Sanchez-Puerta, M. V., Savelkoul, E., Schmitt, I., Slot, J. C., Soanes, D., Szövényi, P., Talbot, N. J., Veneault-Fourrey, C., & Xavier, B. B. (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.

Armaleo, D., Sun, X., & 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.

Armstrong, E., Prost, S., Ertz, D., Westberg, M., Frisch, A., & Bendiksby, M. (2018). Draft genome sequence and annotation of the lichen-forming fungus *Arthonia radiata*. *Genome Announcements*, 6(14), 4–5, doi: 10.1128/genomeA.00281-18.

Asahina, Y. & Shibata, S. (1954). *Chemistry of lichen substances*. Ueno, Tokyo: Japan Society for the Promotion of Science.

Aschenbrenner, I. A., Cernava, T., Berg, G., & Grube, M. (2016). Understanding microbial multi-species symbioses. *Frontiers in Microbiology*, 7, 180, doi: 10.3389/fmicb.2016.00180.

Asplund, J. & Wardle, D. A. (2017). How lichens impact on terrestrial community and ecosystem properties. *Biological Reviews*, 92(3), 1720–1738, doi: 10.1111/brv.12305.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477, doi: 10.1089/cmb.2012.0021.

Barton, D. H. R., Deflorin, A. M., & Edwards, O. E. (1956). 108. The synthesis of usnic acid. *Journal of the Chemical Society*, 108, 530–534, doi: 10.1039/jr9560000530.

Bergmann, S., Schümann, J., Scherlach, K., Lange, C., Brakhage, A. A., & Hertweck, C. (2007). Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nature Chemical Biology*, 3(4), 213–217, doi: 10.1038/nchembio869.

Bertrand, R. L., Abdel-Hameed, M., & Sorensen, J. L. (2018a). 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., Abdel-Hameed, M., & Sorensen, J. L. (2018b). 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.

Bertrand, R. L. & Sorensen, J. L. (2018). A comprehensive catalogue of polyketide synthase gene clusters in lichenizing fungi. *Journal of Industrial Microbiology & Biotechnology*, 45(12), 1067–1081, doi: 10.1007/s10295-018-2080-y.

Bertrand, R. L. & Sorensen, J. L. (2019a). Lost in translation: Challenges with heterologous expression of lichen polyketide synthases. *ChemistrySelect*, 4(21), 6473–6483, doi: 10.1002/slct.201901762.

Bertrand, R. L. & Sorensen, J. L. (2019b). Transcriptional heterologous expression of two type III PKS from the lichen *Cladonia uncialis*. *Mycological Progress*, 18(12), 1437–1447, doi: 10.1007/s11557-019-01539-6.

Bills, G. F. & Gloer, J. B. (2016). Biologically active secondary metabolites from the fungi. *Microbiology Spectrum*, 4(6), FUNK–0009– 2016, doi: 10.1128/microbiolspec.FUNK-0009-2016.

Blin, K., Kim, H. U., Medema, M. H., & Weber, T. (2019). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Briefings in Bioinformatics*, 20(4), 1103–1113, doi: 10.1093/bib/bbx146.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, M. H. (2017). antiSMASH 4.0 – improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, 45(W1), W36–W41, doi: 10.1093/nar/gkx319.

Bode, H. B. & Müller, R. (2005). The impact of bacterial genomics on natural product research. *Angewandte Chemie International Edition*, 44(42), 6828–6846, doi: 10.1002/anie.200501080.

Boettger, D. & Hertweck, C. (2013). Molecular diversity sculpted by fungal PKS-NRPS hybrids. *ChemBioChem*, 14(1), 28–42, doi: 10.1002/cbic.201200624.

Boustie, J. & Grube, M. (2005). Lichens – a promising source of bioactive secondary metabolites. *Plant Genetic Resources: Characterization and Utilization*, 3(2), 273–287, doi: 10.1079/PGR200572.

Brakhage, A. A. (2013). Regulation of fungal secondary metabolism. *Nature Reviews Microbiology*, 11(1), 21–32, doi: 10.1038/nrmicro2916.

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4), 1125–1136, doi: 10.1093/bib/bbx120.

Burlando, B., Ranzato, E., Volante, A., Appendino, G., Pollastro, F., & Verotta, L. (2009). Antiproliferative effects on tumour cells and promotion of keratinocyte wound healing by different lichen compounds. *Planta Medica*, 75(6), 607–613, doi: 10.1055/s-0029-1185329.

Cacho, R. a., Tang, Y., & Chooi, Y.-H. (2015). Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Frontiers in Microbiology*, 5(774), 1–16, doi: 10.3389/fmicb.2014.00774.

Calchera, A., Dal Grande, F., Bode, H. B., & Schmitt, I. (2019). Biosynthetic gene content of the 'perfume lichens' *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules*, 24(1), 203, doi: 10.3390/molecules24010203.

Calcott, M. J., Ackerley, D. F., Knight, A., Keyzers, R. A., & Owen, J. G. (2018). Secondary metabolism in the lichen symbiosis. *Chemical Society Reviews*, 47(5), 1730–1760, doi: 10.1039/C7CS00431A.

Cernava, T., Erlacher, A., Aschenbrenner, I. A., Krug, L., Lassek, C., Riedel, K., Grube, M., & Berg, G. (2017). Deciphering functional diversification within the lichen microbiota by meta-omics. *Microbiome*, 5(1), 82, doi: 10.1186/s40168-017-0303-5.

Chooi, Y.-H. & Solomon, P. S. (2014). A chemical ecogenomics approach to understand the roles of secondary metabolites in fungal cereal pathogens. *Frontiers in Microbiology*, 5(11), 1–7, doi: 10.3389/fmicb.2014.00640.

Chooi, Y.-H., Stalker, D. M., Davis, M. A., Fujii, I., Elix, J. A., Louwhoff, S. H., & Lawrie, A. C. (2008). Cloning and sequence characterization of a non-reducing polyketide synthase gene from the lichen *Xanthoparmelia semiviridis*. *Mycological Research*, 112(2), 147–161, doi: 10.1016/j.mycres.2007.08.022.

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., & Fischbach, M. A. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158(2), 412–421, doi: 10.1016/j.cell.2014.06.034.

Cox, R. J., Skellam, E., & Williams, K. (2018). Biosynthesis of fungal polyketides. In T. Anke & A. Schüffler (Eds.), *Physiology and Genetics. The Mycota (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research)* (pp. 385–412). Cham: Springer, doi: 10.1007/978-3-319-71740-1_13.

Cragg, G. M. & Newman, D. J. (2005). Biodiversity: A continuing source of novel drug leads. *Pure and Applied Chemistry*, 77(1), 7–24, doi: 10.1351/pac200577010007.

Crawford, J. M., Dancy, B. C. R., Hill, E. A., Udwary, D. W., & Townsend, C. A. (2006). Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proceedings of the National Academy of Sciences*, 103(45), 16728–16733, doi: 10.1073/pnas.0604112103.

Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpeh, O., Tsai, S.-C., & Townsend, C. A. (2009). Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature*, 461(7267), 1139–1143, doi: 10.1038/nature08475.

Crawford, J. M., Thomas, P. M., Scheerer, J. R., Vagstad, A. L., Kelleher, N. L., & Townsend, C. A. (2008). Deconstruction of iterative multidomain polyketide synthase function. *Science*, 320(5873), 243–246, doi: 10.1126/science.1154711.

Crawford, J. M. & Townsend, C. A. (2010). New insights into the formation of fungal aromatic polyketides. *Nature Reviews Microbiology*, 8(12), 879–889, doi: 10.1038/nrmicro2465.

Crawford, S. D. (2015). Lichens used in traditional medicine. In B. Ranković (Ed.), *Lichen Secondary Metabolites* chapter 2, (pp. 27–80). Cham: Springer International Publishing, doi: 10.1007/978-3-319-13374-4_2.

Crittenden, P. D., David, J. C., Hawksworth, D. L., & Campbell, F. S. (1995). Attempted isolation and success in the culturing of a broad spectrum of lichen-forming and lichenicolous fungi. *New Phytologist*, 130(2), 267–297, doi: 10.1111/j.1469-8137.1995.tb03048.x.

Culberson, C. F. (1969). *Chemical and botanical guide to lichen products*. Chapel Hill: University of North Carolina Press.

Dal Grande, F., Meiser, A., Greshake Tzovaras, B., Otte, J., Ebersberger, I., & Schmitt, I. (2018a). The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *The Lichenologist*, 50(3), 329–340, doi: 10.1017/S002428291800021X.

Dal Grande, F., Rolshausen, G., Divakar, P. K., Crespo, A., Otte, J., Schleuning, M., & Schmitt, I. (2018b). Environment and host identity structure communities of green algal symbionts in lichens. *New Phytologist*, 217(1), 277–289, doi: 10.1111/nph.14770.

Dal Grande, F., Sharma, R., Meiser, A., Rolshausen, G., Büdel, B., Mishra, B., Thines, M., Otte, J., Pfenninger, M., & Schmitt, I. (2017). Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evolutionary Biology*, 17(1), 93, doi: 10.1186/s12862-017-0929-8.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8), 1164–1165, doi: 10.1093/bioinformatics/btr088.

Davydov, E. A., Peršoh, D., & Rambold, G. (2017). Umbilicariaceae (Lichenized ascomycota) – Trait evolution and a new generic concept. *Taxon*, 66(6), 1282–1303, doi: 10.12705/666.2.

Dias, D. A., Urban, S., & Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites*, 2(2), 303–336, doi: 10.3390/metabo2020303.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K.-S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., & Metcalf, W. W. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology*, 10(11), 963–968, doi: 10.1038/nchembio.1659.

Elix, J. A. (2014). A catalogue of standardized chromatographic data and biosynthetic relationships for lichen substances. *Third Edition. Published by the author, Canberra (Australia)*.

Elix, J. A. & Stocker-Wörgötter, E. (2008). Biochemistry and secondary metabolites. In T. H. Nash (Ed.), *Lichen Biology* chapter 7, (pp. 104–133). Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511790478.008.

Elix, J. A., Whitton, A. A., & Sargent, M. V. (1984). Recent progress in the chemistry of lichen substances. In *Progress in the Chemistry of Organic Natural Products* (pp. 103–234). Vienna: Springer, doi: 10.1007/978-3-7091-8717-3_2.

Erlacher, A., Cernava, T., Cardinale, M., Soh, J., Sensen, C. W., Grube, M., & Berg, G. (2015). *Rhizobiales* as functional and endosymbiontic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Frontiers in Microbiology*, 6(February), 1–9, doi: 10.3389/fmicb.2015.00053.

Fernández-Mendoza, F., Fleischhacker, A., Kopun, T., Grube, M., & Muggia, L. (2017). ITS1 metabarcoding highlights low specificity of lichen mycobiomes at a local scale. *Molecular Ecology*, 26(18), 4811–4830, doi: 10.1111/mec.14244.

Fulton, T. R., Ibrahim, N., Losada, M. C., Grzegorski, D., & Tkacz, J. S. (1999). A melanin polyketide synthase (PKS) gene from *Nodulisporium* sp. that shows homology to the *pks1* gene of *Colletotrichum lagenarium*. *Molecular and General Genetics*, 262(4-5), 714–720, doi: 10.1007/s004380051133.

Gagunashvili, A. N., Davíðsson, S. P., Jónsson, Z. O., & Andrésson, Ó. S. (2009). Cloning and heterologous transcription of a polyketide synthase gene from the lichen *Solorina crocea*. *Mycological Research*, 113(3), 354–363, doi: 10.1016/j.mycres.2008.11.011.

Ghurye, J. S., Cepeda-Espinoza, V., & Pop, M. (2016). Metagenomic assembly: Overview, challenges and applications. *Yale Journal of Biology and Medicine*, 89(3), 353–362.

Goga, M., Elečko, J., Marcinčinová, M., Ručová, D., Bačkorová, M., & Bačkor, M. (2018). Lichen metabolites: An overview of some secondary metabolites and their biological potential. In J. M. Merillon & K. Ramawat (Eds.), *Co-Evolution of Secondary Metabolites* (pp. 1–36). Springer, Cham, doi: 10.1007/978-3-319-76887-8_57-1.

Greshake, B., Zehr, S., Dal Grande, F., Meiser, A., Schmitt, I., & Ebersberger, I. (2016). Potential and pitfalls of eukaryotic metagenome skimming: A test case for lichens. *Molecular Ecology Resources*, 16(2), 511–523, doi: 10.1111/1755-0998.12463.

Greshake Tzovaras, B. (2017). *Characterizing the hologenome of Lasallia pustulata and tracing genomic footprints of lichenization*. PhD thesis, Johann Wolfgang Goethe University, Frankfurt am Main, Germany.

Greshake Tzovaras, B., Segers, F. H. I. D., Bicker, A., Dal Grande, F., Otte, J., Anvar, S. Y., Hankeln, T., Schmitt, I., & Ebersberger, I. (2019). What is in a lichen? A metagenomic approach to reconstruct the holo-genome of *Umbilicaria pustulata*. *bioRxiv*, (pp. 1–36)., doi: 10.1101/810986.

Grube, M., Berg, G., S. Andrésson, Ó., Vilhelmsson, O., Dyer, P. S., & Miao, V. P. (2013). Lichen genomics. In F. Martin (Ed.), *The ecological genomics of fungi* chapter 9, (pp. 191–212). Hoboken, NJ: John Wiley & Sons, Inc, doi: 10.1002/9781118735893.ch9.

Grube, M. & Wedin, M. (2016). Lichenized fungi and the evolution of symbiotic organization. *Microbiology Spectrum*, 4(6), 1–17, doi: 10.1128/microbiolspec.FUNK-0011-2016.

Haider, B., Ahn, T.-H. H., Bushnell, B., Chai, J., Copeland, A., & Pan, C. (2014). Omega: an overlap-graph de novo assembler for metagenomics. *Bioinformatics*, 30(19), 2717–2722, doi: 10.1093/bioinformatics/btu395.

Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery*, 14(2), 111–129, doi: 10.1038/nrd4510.

Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angewandte Chemie*, 48(26), 4688–4716, doi: 10.1002/anie.200806121.

Hestmark, G. (1992). Sex, size, competition and escape - strategies of reproduction and dispersal in Lasallia pustulata (Umbilicariaceae, Ascomycetes). *Oecologia*, 92, 305–312.

Huneck, S. (1999). The significance of lichens and their metabolites. *Naturwissenschaften*, 86(12), 559–570.

Huneck, S. & Yoshimura, I. (1996). *Identification of lichen substances*. Berlin, Heidelberg: Springer-Verlag.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., Pace, N., Stahl, D., Lane, D., Olsen, G., Handelsman, J., Rondon, M., Brady, S., Clardy, J., Goodman, R., Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J., Wheeler, D., Mitchell, A., Chang, H., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., Mackelprang, R., Waldrop, M., DeAngelis, K., David, M., Chavarria, K., Blazewicz, S., Kanehisa, M., Goto, S., Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., Jansson, J., Buchfink, B., Xie, C., Huson, D., Huson, D., Mitra, S., Weber, N., Ruscheweyh, H., Schuster, S., Huson, D., Auch, A., Qi, J., Schuster, S., Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., Overbeek, R., Olson, R., Pusch, G., Olsen, G., Davis, J., Disz, T., Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Willmann, M., El-Hadidi, M., Huson, D., Schütz, M., Weidenmaier, C., Autenrieth, I., Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., Wood, D., Salzberg, S., Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., Glass, E., Wilkening, J., Wilke, A., Antonopoulos, D., Meyer, F., Peabody, M., Rossum, T. V., Lo, R., Brinkman, F., Lindgreen, S., Adair, K., Gardner, P., Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Servick, K., Buchfink, B., Huson, D., Xie, C., Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Goodrich, J., Rienzi, S., Poole, A., Koren, O., Walters, W., Caporaso, J., Bray, R., Curtis, J., Treangen, T., Koren, S., Sommer, D., Liu, B., Astrovskaya, I., Ondov, B., Howe, A., Jansson, J., Malfatti, S., Tringe, S., Tiedje, J., Brown, C., Wang, Q., Fish, J., Gilman, M., Sun, Y., Brown, C., & Tiedje, J. (2016). MEGAN community edition – Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, 12(6), 4–12, doi: 10.1371/journal.pcbi.1004957.

Jenke-Kodama, H., Sandmann, A., Müller, R., & Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution*, 22(10), 2027–2039, doi: 10.1093/molbev/msi193.

Joulain, D. & Tabacchi, R. (2009a). Lichen extracts as raw materials in perfumery. Part 1: oakmoss. *Flavour and Fragrance Journal*, 24(2), 49–61, doi: 10.1002/ffj.1916.

Joulain, D. & Tabacchi, R. (2009b). Lichen extracts as raw materials in perfumery. Part 2: treemoss. *Flavour and Fragrance Journal*, 24(3), 105–116, doi: 10.1002/ffj.1923.

Junttila, S. & Rudd, S. (2012). Characterization of a transcriptome from a non-model organism, *Cladonia rangiferina*, the grey reindeer lichen, using high-throughput next generation sequencing and EST sequence data. *BMC Genomics*, 13(1), 575, doi: 10.1186/1471-2164-13-575.

Kampa, A., Gagunashvili, A. N., Gulder, T. A. M., Morinaka, B. I., Daolio, C., Godejohann, M., Miao, V. P. W., Piel, J., & Andrésson, Ó. S. (2013). Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proceedings of the National Academy of Sciences*, 110(33), E3129–E3137, doi: 10.1073/pnas.1305867110.

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066, doi: 10.1093/nar/gkf436.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780, doi: 10.1093/molbev/mst010.

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., van Santen, J. A., Tracanna, V., Suarez Duran, H. G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian, L. K., Collemare, J., Linington, R. G., Weber, T., & Medema, M. H. (2019). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*, gkz882, 1–5, doi: 10.1093/nar/gkz882.

Keller, N. P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. *Nature Reviews Microbiology*, 17(3), 167–180, doi: 10.1038/s41579-018-0121-1.

Keller, N. P., Turner, G., & Bennett, J. W. (2005). Fungal secondary metabolism – from biochemistry to genomics. *Nature Reviews Microbiology*, 3(12), 937–947, doi: 10.1038/nrmicro1286.

Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C., & Turgeon, B. G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences*, 100(26), 15670–15675, doi: 10.1073/pnas.2532165100.

Kumar KC, S. & Müller, K. (1999). Lichen metabolites. 2. Antiproliferative and cytotoxic activity of gyrophoric, usnic, and diffractaic acid on human keratinocyte growth. *Journal of Natural Products*, 62(6), 821–823, doi: 10.1021/np980378z.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 5(R12), 1–9, doi: 10.1186/gb-2004-5-2-r12.

Lantz, H., Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J. F., Vlasova, A., Leskosek, B. L., Soler, L., & Binzer-Panchal, M. (2018). Ten steps to get started in genome assembly and annotation. *F1000Research*, 7, 1–25, doi: 10.12688/f1000research.13598.1.

Lawrey, J. D. (1986). Biological role of lichen substances. *The Bryologist*, 89(2), 111, doi: 10.2307/3242751.

Lendemer, J. C., Keepers, K. G., Tripp, E. A., Pogoda, C. S., McCain, C. M., & Kane, N. C. (2019). A taxonomically broad metagenomic survey of 339 species spanning 57 families suggests cystobasidiomycete yeasts are not ubiquitous across all lichens. *American Journal of Botany*, 106(8), 1090–1095, doi: 10.1002/ajb2.1339.

Lewis, K. (2013). Platforms for antibiotic discovery. *Nature Reviews Drug Discovery*, 12(5), 371–387, doi: 10.1038/nrd3975.

Lindgreen, S., Adair, K. L., & Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, 6(1), 19233, doi: 10.1038/srep19233.

Lücking, R., Hodkinson, B. P., & Leavitt, S. D. (2016). 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.

Lumbsch, H. T. (1998). Taxonomic use of metabolic data in lichen-forming fungi. In J. C. Frisvad, P. D. Bridge, & D. Arora (Eds.), *Chemical fungal taxonomy* chapter 14, (pp. 345–387). New York, U.S.: Marcel Dekker, Inc.

Machado, H., Tuttle, R. N., & Jensen, P. R. (2017). Omics-based natural product discovery and the lexicon of genome mining. *Current Opinion in Microbiology*, 39, 136–142, doi: 10.1016/j.mib.2017.10.025.

McDonald, T. R., Gaya, E., & Lutzoni, F. (2013a). 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.

McDonald, T. R., Mueller, O., Dietrich, F. S., & Lutzoni, F. (2013b). High-throughput genome sequencing of lichenizing fungi to assess gene loss in the ammonium transporter/ammonia permease gene family. *BMC Genomics*, 14(1), 225, doi: 10.1186/1471-2164-14-225.

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J. N., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kötter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gómez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J. M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Müller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Süssmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., & Glöckner, F. O. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology*, 11(9), 625–631, doi: 10.1038/nchembio.1890.

Meiser, A., Otte, J., Schmitt, I., & Dal Grande, F. (2017). Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi. *Scientific Reports*, 7(1), 14881, doi: 10.1038/s41598-017-14576-6.

Miao, V. P., Manoharan, S. S., Snæbjarnarson, V., & Andrésson, Ó. S. (2012). Expression of *lec-1*, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea*. *Symbiosis*, 57(1), 23–31, doi: 10.1007/s13199-012-0175-1.

Miao, V. P. W., Coëffet-LeGal, M.-F., Brown, D. W., Sinnemann, S. J., Donaldson, G., & Davies, J. (2001). Genetic approaches to harvesting lichen products. *Trends in Biotechnology*, 19(9), 349–355, doi: 10.1016/S0167-7799(01)01700-0.

Mishra, B. B. & Tiwari, V. K. (2011). Natural products: An evolving role in future drug discovery. *European Journal of Medicinal Chemistry*, 46(10), 4769–4807, doi: 10.1016/j.ejmech.2011.07.057.

Molnár, K. & 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.

Moreno-Hagelsieb, G. & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, 24(3), 319–324, doi: 10.1093/bioinformatics/btm585.

Mouncey, N. J., Otani, H., Udwary, D., & Yoshikuni, Y. (2019). New voyages to explore the natural product galaxy. *Journal of Industrial Microbiology and Biotechnology*, 46(3), 273–279, doi: 10.1007/s10295-018-02122-w.

Moya, P., Molins, A., Martínez-Alberola, F., Muggia, L., & Barreno, E. (2017). Unexpected associated microalgal diversity in the lichen *Ramalina farinacea* is uncovered by pyrosequencing analyses. *PLoS ONE*, 12(4), e0175091, doi: 10.1371/journal.pone.0175091.

Muggia, L. & Grube, M. (2018). Fungal diversity in lichens: From extremotolerance to interactions with algae. *Life*, 8(2), 15, doi: 10.3390/life8020015.

Muggia, L., Schmitt, I., & Grube, M. (2009). Lichens as treasure chests of natural products. *Sim News*, 59, 85–97.

Nagarajan, N. & Pop, M. (2013). Sequence assembly demystified. *Nature Reviews Genetics*, 14(3), 157–67, doi: 10.1038/nrg3367.

Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research*, 40(20), e155, doi: 10.1093/nar/gks678.

Nash, T. H. (2008). *Lichen biology*. Cambridge: Cambridge University Press, 2 edition, doi: 10.1017/CBO9780511790478.

Nasko, D. J., Koren, S., Phillippy, A. M., & Treangen, T. J. (2018). RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. *Genome Biology*, 19(1), 165, doi: 10.1186/s13059-018-1554-6.

Newman, D. J. & Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products*, 79(3), 629–661, doi: 10.1021/acs.jnatprod.5b01055.

Nielsen, J. C., Grijseels, S., Prigent, S., Ji, B., Dainat, J., Nielsen, K. F., Frisvad, J. C., Workman, M., & Nielsen, J. (2017). Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nature Microbiology*, 2, 17044, doi: 10.1038/nmicrobiol.2017.44.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834, doi: 10.1101/gr.213959.116.

Park, S.-Y., Choi, J., Kim, J. A., Jeong, M.-H., Kim, S., Lee, Y.-H., & Hur, J.-S. (2013a). Draft genome sequence of *Cladonia macilenta* KoLRI003786, a lichen-forming fungus producing biruloquinone. *Genome Announcements*, 1(5), e00695–13, doi: 10.1128/genomeA.00695-13.

Park, S.-Y., Choi, J., Kim, J. A., Yu, N.-H., Kim, S., Kondratyuk, S. Y., Lee, Y.-H., & Hur, J.-S. (2013b). Draft genome sequence of lichen-forming fungus *Caloplaca flavorubescens* strain KoLRI002931. *Genome Announcements*, 1(4), e00678–13, doi: 10.1128/genomeA.00678-13.

Park, S.-Y., Choi, J., Lee, G.-W., Jeong, M.-H., Kim, J. A., Oh, S.-O., Lee, Y.-H., & Hur, J.-S. (2014a). Draft genome sequence of *Umbilicaria muehlenbergii* KoLRILF000956, a lichen-forming fungus amenable to genetic manipulation. *Genome Announcements*, 2(2), e00357–14, doi: 10.1128/genomeA.00357-14.

Park, S.-Y., Choi, J., Lee, G.-W., Kim, J. A., Oh, S.-O., Jeong, M.-H., Yu, N.-H., Kim, S., Lee, Y.-H., & Hur, J.-S. (2014b). Draft genome sequence of lichen-forming fungus *Cladonia metacorallifera* strain KoLRI002260. *Genome Announcements*, 2(1), e01065–13, doi: 10.1128/genomeA.01065-13.

Park, S.-Y., Choi, J., Lee, G.-W., Park, C.-H., Kim, J. A., Oh, S.-O., Lee, Y.-H., & Hur, J.-S. (2014c). Draft genome sequence of *Endocarpon pusillum* strain KoLRILF000583. *Genome Announcements*, 2(3), e00452–14, doi: 10.1128/genomeA.00452-14.

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420–1428, doi: 10.1093/bioinformatics/bts174.

Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics*, 10(4), 354–366, doi: 10.1093/bib/bbp026.

Posner, B., Feige, G. B., & Leuckert, C. (1991). Beiträge zur Chemie der Flechtengattung *Lasallia* Mérat / On the Chemistry of the lichen genus *Lasallia* Mérat. *Zeitschrift für Naturforschung C*, 46(1-2), 19–27, doi: 10.1515/znc-1991-1-204.

Ranković, B. & Kosanić, M. (2015). *Lichen secondary metabolites*. Cham: Springer International Publishing, doi: 10.1007/978-3-319-13374-4.

Rokas, A., Wisecaver, J. H., & 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.

Rundel, P. W. (1978). The ecological role of secondary lichen substances. *Biochemical Systematics and Ecology*, 6(3), 157–170, doi: 10.1016/0305-1978(78)90002-9.

Sanchez, J. F., Chiang, Y.-M., Szewczyk, E., Davidson, A. D., Ahuja, M., Elizabeth Oakley, C., Woo Bok, J., Keller, N., Oakley, B. R., & Wang, C. C. C. (2010). Molecular genetic analysis of the orsellinic acid/F9775 genecluster of *Aspergillus nidulans*. *Molecular BioSystems*, 6(3), 587–593, doi: 10.1039/B904541D.

Sancho, L. G. & Crespo, A. (1989). *Lasallia hispanica* and related species. *The Lichenologist*, 21(1), 45–58, doi: 10.1017/S002428298900006X.

Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1), 8, doi: 10.1186/s40168-016-0154-5.

Schmitt, I., Martín, M. P., Kautz, S., & Lumbsch, H. T. (2005). Diversity of non-reducing polyketide synthase genes in the Pertusariales (lichenized Ascomycota): A phylogenetic perspective. *Phytochemistry*, 66(11), 1241–1253, doi: 10.1016/j.phytochem.2005.04.014.

Schümann, J. & Hertweck, C. (2006). Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *Journal of Biotechnology*, 124(4), 690–703, doi: 10.1016/j.jbiotec.2006.03.046.

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., DeMaere, M. Z., Chikhi, R., Nagarajan, N., Quince, C., Meyer, F., Balvočiūtė, M., Hansen, L. H., Sørensen, S. J., Chia, B. K. H., Denis, B., Froula, J. L., Wang, Z., Egan, R., Don Kang, D., Cook, J. J., Deltel, C., Beckstette, M., Lemaitre, C., Peterlongo, P., Rizk, G., Lavenier, D., Wu, Y.-W., Singer, S. W., Jain, C., Strous, M., Klingenberg, H., Meinicke, P., Barton, M. D., Lingner, T., Lin, H.-H., Liao, Y.-C., Silva, G. G. Z., Cuevas, D. A., Edwards, R. A., Saha, S., Piro, V. C., Renard, B. Y., Pop, M., Klenk, H.-P., Göker, M., Kyrpides, N. C., Woyke, T., Vorholt, J. A., Schulze-Lefert, P., Rubin, E. M., Darling, A. E., Rattei, T., & McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Methods*, 14(11), 1063–1071, doi: 10.1038/nmeth.4458.

Seaward, M. R. D. (2008). Environmental role of lichens. In T. H. Nash (Ed.), *Lichen Biology* chapter 14, (pp. 274–298). Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511790478.015.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212, doi: 10.1093/bioinformatics/btv351.

Soderlund, C., Bomhoff, M., & Nelson, W. M. (2011). SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Research*, 39(10), e68–e68, doi: 10.1093/nar/gkr123.

Spribille, T. (2018). Relative symbiont input and the lichen symbiotic outcome. *Current Opinion in Plant Biology*, 44, 57–63, doi: 10.1016/j.pbi.2018.02.007.

Spribille, T., Tuovinen, V., Resl, P., Vanderpool, D., Wolinski, H., Aime, M. C., Schneider, K., Stabentheiner, E., Toome-Heller, M., Thor, G., Mayrhofer, H., Johannesson, H., & McCutcheon, J. P. (2016). Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science*, 353(6298), 488–492, doi: 10.1126/science.aaf8287.

Stocker-Wörgötter, E. (2008). Metabolic diversity of lichen-forming ascomycetous fungi: culturing, polyketide and shikimatemetabolite production, and PKS genes. *Natural Product Reports*, 25(1), 188–200, doi: 10.1039/B606983P.

Stocker-Wörgötter, E. (2015). Biochemical diversity and ecology of lichen-forming fungi: Lichen substances, chemosyndromic variation and origin of polyketide-type metabolites (biosynthetic pathways). In D. K. Upreti, P. K. Divakar, V. Shukla, & R. Bajpai (Eds.), *Recent Advances in Lichenology* chapter 9, (pp. 161–179). New Delhi: Springer India, doi: 10.1007/978-81-322-2235-4_9.

Strous, M., Kraft, B., Bisdorf, R., & Tegetmeyer, H. E. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in Microbiology*, 3(12), 1–11, doi: 10.3389/fmicb.2012.00410.

Taguchi, H., Sankawa, U., & Shibata, S. (1969). Biosynthesis of natural products. VI. Biosynthesis of usnic acid in lichens. (1). A general scheme of biosynthesis of usnic acid. *Chemical & Pharmaceutical Bulletin*, 17(10), 2054–2060, doi: 10.1248/cpb.17.2054.

Tringe, S. G. & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, 6(11), 805–814, doi: 10.1038/nrg1709.

Tuovinen, V., Ekman, S., Thor, G., Vanderpool, D., Spribille, T., & Johannesson, H. (2019). Two basidiomycete fungi in the cortex of wolf lichens. *Current Biology*, 29(3), 476–483.e5, doi: 10.1016/j.cub.2018.12.022.

Vivas, M., Pérez-Ortega, S., Pintado, A., & Sancho, L. G. (2017). *Fv/Fm* acclimation to the Mediterranean summer drought in two sympatric *Lasallia* species from the Iberian mountains. *The Lichenologist*, 49(2), 157–165, doi: 10.1017/S0024282917000032.

Vogel, K. J. & Moran, N. A. (2013). Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biology and Evolution*, 5(5), 891–904, doi: 10.1093/gbe/evt054.

Vollmers, J., Wiegand, S., & Kaster, A.-K. (2017). Comparing and evaluating metagenome assembly tools from a microbiologist's perspective – not only size matters! *PLoS ONE*, 12(1), e0169662, doi: 10.1371/journal.pone.0169662.

Wang, X., Zhang, X., Liu, L., Xiang, M., Wang, W., Sun, X., Che, Y., Guo, L., Liu, G., Guo, L., Wang, C., Yin, W.-B., Stadler, M., Zhang, X., & Liu, X. (2015). Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. *BMC Genomics*, 16(1), 28, doi: 10.1186/s12864-014-1190-9.

Wang, Y., Geng, C., Yuan, X., Hua, M., Tian, F., & Li, C. (2018a). Identification of a putative polyketide synthase gene involved in usnic acid biosynthesis in the lichen *Nephromopsis pallescens*. *PLoS ONE*, 13(7), e0199110, doi: 10.1371/journal.pone.0199110.

Wang, Y., Neng, Z., Xiaolong, Y., Mei, H., Jae-Seoun, H., Yang, Y., & Wang, J. (2016). Heterologous transcription of a polyketide synthase gene from the lichen forming fungi *Usnea longissima*. *Research Journal of Biotechnology*, 11(10), 16–21.

Wang, Y., Wang, J., Cheong, Y. H., & Hur, J.-S. (2014a). Three new non-reducing polyketide synthase genes from the lichen-forming fungus *Usnea longissima*. *Mycobiology*, 42(1), 34–40, doi: 10.5941/MYCO.2014.42.1.34.

Wang, Y., Yuan, X., Chen, L., Wang, X., & Li, C. (2018b). Draft genome sequence of the lichen-forming fungus *Ramalina intermedia* Strain YAF0013. *Genome Announcements*, 6(23), 4–6, doi: 10.1128/genomeA.00478-18.

Wang, Y.-Y., Liu, B., Zhang, X.-Y., Zhou, Q.-M., Zhang, T., Li, H., Yu, Y.-F., Zhang, X.-L., Hao, X.-Y., Wang, M., Wang, L., & Wei, J.-C. (2014b). Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genomics*, 15(1), 34, doi: 10.1186/1471-2164-15-34.

Ward, N. & Moreno-Hagelsieb, G. (2014). Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE*, 9(7), e101850, doi: 10.1371/journal.pone.0101850.

Werth, S., Miao, V. P. W., Jónsson, Z. O., & Andrésson, Ó. S. (2015). High-throughput sequencing in studies of lichen population biology. In D. K. Upreti, P. K. Divakar, V. Shukla, & R. Bajpai (Eds.), *Recent Advances in Lichenology* chapter 4, (pp. 61–94). New Delhi: Springer, doi: 10.1007/978-81-322-2235-4_4.

Wilson, M. C., Mori, T., Rückert, C., Uria, A. R., Helf, M. J., Takada, K., Gernert, C., Steffens, U. A. E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E. J. N., Brachmann, A. O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., & Piel, J. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature*, 506(7486), 58–62, doi: 10.1038/nature12959.

Wogan, G. N. (1966). Chemical nature and biological effects of the aflatoxins. *Bacteriological Reviews*, 30(2), 460–470.

Wood, D. E. & 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.

Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology*, 6(2), e1000667, doi: 10.1371/journal.pcbi.1000667.

Wright, G. D. (2019). Unlocking the potential of natural products in drug discovery. *Microbial Biotechnology*, 12(1), 55–57, doi: 10.1111/1751-7915.13351.

Xavier, B. B., Miao, V. P. W., Jónsson, Z. O., & Andrésson, Ó. S. (2012). Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: features and phylogeny. *Fungal Biology*, 116(7), 802–814, doi: 10.1016/j.funbio.2012.04.013.

Zhang, A., Lu, P., Dahl-Roshak, A. M., Paress, P. S., Kennedy, S., Tkacz, J. S., & An, Z. (2003). Efficient disruption of a polyketide synthase gene (*pks1*) required for melanin synthesis through *Agrobacterium*-mediated transformation of *Glarea lozoyensis*. *Molecular Genetics and Genomics*, 268(5), 645–55, doi: 10.1007/s00438-002-0780-4.

Zhang, Q. (2018). Metagenome assembly and contig assignment. *Methods in Molecular Biology*, 1849, 179–192, doi: 10.1007/978-1-4939-8728-3_12.

Zhao, H. & Medema, M. H. (2016). Standardization for natural product synthetic biology. *Natural Product Reports*, 33(8), 920–924, doi: 10.1039/C6NP00030D.

Zhou, R., Yang, Y., Park, S.-Y., Nguyen, T. T., Seo, Y.-W., Lee, K. H., Lee, J. H., Kim, K. K., Hur, J.-S., & Kim, H. (2017). The lichen secondary metabolite atranorin suppresses lung cancer cell motility and tumorigenesis. *Scientific Reports*, 7(1), 8136, doi: 10.1038/s41598-017-08225-1.

Ziemert, N. & Jensen, P. R. (2012). Phylogenetic approaches to natural product structure prediction. In *Methods in Enzymology*, volume 517 chapter 8, (pp. 161–182). Elsevier Inc., 1 edition, doi: 10.1016/B978-0-12-404634-4.00008-5.

Zopf, W. F. (1907). *Die Flechtenstoffe in chemischer, botanischer, pharmakologischer und technischer Beziehung*. Jena: Fischer.

Glossary

A	adenylation domain
ACP	acyl-carrier protein
AT	acyltransferase domain
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	base pairs
C	condensation domain
cDNA	complementary deoxyribonucleic acid
DH	dehydratase domain
DNA	deoxyribonucleic acid
E	epimerization domain
ER	enoyl reductase domain
GO	gene ontology
kb	kilo base pairs
KR	β-ketoacyl reductase domain
KS	ketoacyl synthase domain
LCA	lowest common ancestor
Mb	Mega base pairs
mRNA	messenger ribonucleic acid
MT	methyltransferase domain
N50	minimum contig length needed to cover 50 % of the genome
NCBI	National Center for Biotechnology Information
NR-PKS	non-reducing polyketide synthase
NRPS	non-ribosomal peptide synthetase
PCP	peptidyl-carrier protein domain (also known as thiolation domain)
PKS	polyketide synthase
PKS-NRPS	hybrid polyketide synthase – non-ribosomal peptide synthetase
PPT	4'-phosphopantetheine transferase domain
PT	product template domain
R-PKS	partially or highly reducing polyketide synthase
RBH	Reciprocal Best Blast Hits
RNA	ribonucleic acid
SAT	starter acyl carrier protein transacylase domain
TE	thioesterase domain
UV	ultraviolet

A. Appendix: Publications

A.1. Metagenomic reconstruction of lichen-forming fungi

Statement of author contributions

Title: Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi

Journal: Scientific Reports 7(1), 14881, doi:10.1038/s41598-017-14576-6

Status: Accepted on 12 October 2017

Authors: Anjuli Meiser (AM), Jürgen Otte (JO), Imke Schmitt (IS), Francesco Dal Grande (FDG)

Author contributions:

(1) Development and planning	
Doctoral candidate (AM):	50 %
Coautor FDG:	50 %

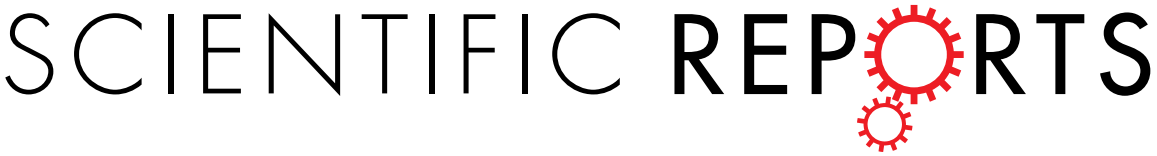
(2) Conduction of experiments	
Coautor IS:	25 % Sample collection from the field
Coautor FDG:	25 % Sample collection from the field
Coautor JO:	50 % Cultivation and DNA/RNA isolation

(3) Compilation of data sets and figures/tables	
Doctoral candidate (AM):	80 % Data quality trimming, genome assemblies, gene prediction, annotation, taxonomic assignments, preparation of all figures and tables
Coautor FDG:	20 % Help with processing the data

(4) Data analyses and interpretation	
Doctoral candidate (AM):	80 % All data analyses (Assembler and taxonomic assignment evaluation, comparison of gene sets, genome mining results), data interpretations
Coautor FDG:	10 % Data interpretation
Coautor IS:	10 % Data interpretation

(5) Preparation of manuscript	
Doctoral candidate (AM):	70 %
Coautor FDG:	20 %
Coautor IS:	10 %

www.nature.com/scientificreports



OPEN

Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi

Received: 15 June 2017
Accepted: 12 October 2017
Published online: 02 November 2017

Anjuli Meiser[1,2], Jürgen Otte[2], Imke Schmitt[1,2] & Francesco Dal Grande[2]

The metagenome skimming approach, i.e. low coverage shotgun sequencing of multi-species assemblages and subsequent reconstruction of individual genomes, is increasingly used for in-depth genomic characterization of ecological communities. This approach is a promising tool for reconstructing genomes of facultative symbionts, such as lichen-forming fungi, from metagenomic reads. However, no study has so far tested accuracy and completeness of assemblies based on metagenomic sequences compared to assemblies based on pure culture strains of lichenized fungi. Here we assembled the genomes of *Evernia prunastri* and *Pseudevernia furfuracea* based on metagenomic sequences derived from whole lichen thalli. We extracted fungal contigs using two different taxonomic binning methods, and performed gene prediction on the fungal contig subsets. We then assessed quality and completeness of the metagenome-based assemblies using genome assemblies as reference which are based on pure culture strains of the two fungal species. Our comparison showed that we were able to reconstruct fungal genomes from uncultured lichen thalli, and also cover most of the gene space (86–90%). Metagenome skimming will facilitate genome mining, comparative (phylo)genomics, and population genetics of lichen-forming fungi by circumventing the time-consuming, sometimes unfeasible, step of aposymbiotic cultivation.

In recent years, the decreasing costs and higher accessibility of high-throughput DNA sequencing technologies have revolutionized microbial ecology research. Direct sequencing of genomic material from the environment, commonly referred to as metagenomics, can provide a cultivation-independent assessment of the largely untapped genetic diversity and functional aspects of microbial communities. Whole-metagenome shotgun sequencing has been applied to study diverse microbiomes, spanning a range of natural environments, including the human body[1–3]. Metagenomics has not only been used to catalogue diversity, but it has also provided a fresh perspective on our understanding of the intricate, multi-species interactions driving symbiotic communities, and how these interactions influence ecosystems[4]. On the other hand, the conversion of these large volumes of sequencing data to biologically useful information remains a major challenge[5].

With the improvement of bioinformatics tools, it is increasingly possible to assemble whole genomes from environmental communities of both prokaryotes and eukaryotes, and analyse their strain-level variation[6]. Although research on metagenomic assembly is still in its infancy, valuable insights have already been derived[7]. The annotation of metagenomic contigs from multi-species communities has proven useful to study evolutionary patterns, metabolic complementation, genetic exchange and/or modification between symbionts and their hosts in several symbiotic systems. The reconstruction of individual genomes from multi-species communities has also been used to isolate genes associated with the biosynthesis of novel biomolecules[8]. Assembly and annotation of sequencing data, however, pose several analytical challenges[9]. In particular, the co-occurrence of multiple strains or similar species – sometimes present at highly uneven ratios – may drastically reduce the quality of the reconstructed genomes[10].

[1]Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt, Max-von-Laue Str. 13, D-60438, Frankfurt, Germany. [2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, D-60486, Frankfurt, Germany. Correspondence and requests for materials should be addressed to I.S. (email: imke.schmitt@senckenberg.de) or F.D.G. (email: francesco.dalgrande@senckenberg.de)

The lichen symbiosis is an example of a multi-species symbiotic assemblage, which we begin to understand much better since the advent of next-generation sequencing technologies[11]. In fact lichens are not simply an obligate association between a fungal (mycobiont) and a photosynthetic partner (photobiont), which can be either a cyanobacterium and/or a green alga[12]. The long-lived thalli of lichens constitute microhabitats harbouring a surprisingly high diversity of other eukaryotic and prokaryotic (both bacteria and archaea) microorganisms whose function has not yet been established[13,14]. Furthermore, next-generation sequencing data revealed the large extent to which multiple fungal species, and multiple photobiont lineages can be present within the same lichen individual[15–17]. Lichen-forming fungi are also relevant in natural product research as they produce a vast array of natural compounds many of which are bioactive[18–21]. To tackle evolutionary, ecological and biotechnological aspects of the lichen symbiosis, researchers have begun implementing metagenomic tools[22–26]. This is particularly relevant for studying lichen-forming fungi because these fungi are tedious to isolate[27]. Aposymbiotic cultivation of lichenized fungi is impeded by i) unknown culture conditions, ii) external fungal and bacterial contamination, iii) slow growth rates. Further, due to the obligate nature of the lichen symbiosis, for many lichen-forming fungi aposymbiotic culturing might not be possible at all[28]. For these reasons, metagenomic tools represent a promising, culture-independent approach to obtain genetic information on the lichen-forming fungi. However, we know little about the challenges and potential biases affecting the genomic assembly of metagenomic reads from whole lichen thalli.

The use of a single sequencing library layout ("metagenome skimming") has been proven a viable approach to reconstruct genomes of the individual lichen symbiotic components, in particular the fungus[22,29]. Two approaches have been implemented, i) sequencing putative fungal DNA from portions of thalli from which algal and other possible contaminants had been manually removed[30], ii) sequencing DNA isolated from whole thalli and extracting putative fungal contigs bioinformatically[29]. The first approach is not always feasible, because the morphology of many species precludes the physical separation of fungal and algal (and other potentially contaminating) cells. The second approach has two main disadvantages: i) assembly strategy depends on the individual experimental set-up, and particular attention should be paid to data with extreme coverage biases, ii) quality of the resulting fungal contig set depends on the assignment method, and on the database used for taxonomic assignment. As the number of studies utilising the metagenome skimming approach is destined to increase in the future, it is important to evaluate accuracy, completeness, and reliability of the method in reconstructing fungal genomes from whole lichen thalli.

Here we assessed the general applicability of the metagenome skimming approach for reconstructing the genome of lichen-forming fungi from whole thalli. For this purpose, we compared fungal assemblies extracted from metagenomic contig sets with the genomes obtained from pure cultures of the respective fungal species. As study systems we chose two lichens, *Evernia prunastri* (also known as oak moss) and *Pseudevernia furfuracea* (also known as tree moss), which are used in the fragrance industry[31,32]. Specifically, we addressed the following research questions: (i) Can metagenome assemblies be used to retrieve the fungal genome and gene space of a lichen-forming fungus? (ii) To what extent is contamination affecting our ability to reconstruct genomes of lichenized fungi from metagenomic samples?

Material and Methods

Fungal cultures and genome sequencing. The culture of the lichen-forming fungus *P. furfuracea* was obtained from the AKITA culture collection (collection number 0122 M). The culture of the lichen-forming fungus *E. prunastri* was obtained by picking single vegetative hyphal cells from a squash preparation of a lichen thallus using a micromanipulator following the protocol by Beck & Koop[33]. Details of the materials are given in Table 1. We grew the two fungal cultures on malt yeast extract medium. Cultures were kept in darkness in a climate chamber at 16 °C. We sub-cultured every two to three months onto fresh medium until sufficient biomass (~1 g) for genome sequencing was obtained.

We isolated genomic DNA from each mycobiont culture following the CTAB Maxi-prep method[34] after grinding the mycelium in liquid nitrogen with a mortar and pestle. The DNA was further purified with the PowerClean DNA Clean-Up Kit (MO BIO, Carlsbad, CA, USA) and sequenced using different platforms and library layouts. For the culture of *E. prunastri* we sequenced the following libraries: 300 bp paired-end library, 800 bp paired-end library and 3 kbp mate-pair library, on Illumina HiSeq (100 bp × 2). For the culture of *P. furfuracea* we sequenced the following libraries: 300 bp paired-end library on Illumina MiSeq (300 bp × 2) and two mate-pair libraries (3 kbp and 8 kbp) on Illumina HiSeq (150 bp × 2).

Sequencing of metagenomes and metatranscriptomes. For the metagenomes, we sequenced genomic DNA isolated from whole lichen thalli of *E. prunastri* and *P. furfuracea* (one thallus each). Voucher information is given in Table 1. We washed the thalli thoroughly with sterile water, and checked under the stereomicroscope that thalli were free from visible parasitic infections. We isolated and purified genomic DNA as described above. For metagenome sequencing we chose a single library layout (250 bp × 2 Illumina MiSeq).

Additionally, we sequenced the metatranscriptome of *E. prunastri* and *P. furfuracea* to provide RNA-based evidence for improving gene model prediction. For *P. furfuracea* we isolated RNA from both chemical variants (i.e., chemotypes) of the species, the olivetoric acid and the physodic acid chemotypes. Whole lichen thalli were collected and stored directly in RNAlater (Sigma-Aldrich Chemie GmbH, Munich, Germany) (Table 1). Total RNA was isolated using the method described by Rubio-Piña & Zapata-Pérez[35] after blotting the thalli dry and grinding them in liquid nitrogen with a mortar and pestle. The isolated poly-A^+ RNA was further purified with the RNeasy MinElute Clean-up Kit (Qiagen, Hilden, Germany), and sequenced on Illumina MiSeq at StarSeq (Mainz, Germany) and a 250 bp paired-end library for both *P. furfuracea* chemotypes, and a 300 bp paired-end library for *E. prunastri*.

Species	Type of data generated	Source of DNA	Voucher information	Herbarium/culture code	NCBI accession number
Evernia prunastri	genome	fungal culture	Spain, 28048 Madrid, Fuencarral-El Pardo, 621 m; 40.48822, −3.75026; leg. F. Dal Grande & P. K. Divakar June, 2012	Imke Schmitt lab, SB iK-F, C 0001	NKYR00000000
Evernia prunastri	metagenome	thallus	Norway, Jeløya, 1519 Moss, 19 m; 59.42553, 10.60794; | eg. F. Dal Grande & G. Singh, August 2012	FR-0265082	SRS2339650
Evernia prunastri	metatranscriptome	thallus	Germany, Hesse, 60388 Frankfurt/Main, Bornweidstraße 42, 102 m; 50.148683, 8.758133; leg. I. Schmitt, January 2014	FR-0265083	SRS2339648
Pseudevernia furfuracea	genome	fungal culture	Slovenia, 17.10.1996, leg. Isao Yoshimura, originated from thallus; olivetoric acid chemotype	AKPM 0122M	NKYQ00000000
Pseudevernia furfuracea	metagenome	thallus	Germany, Hesse, Taunus, Großer Feldberg, 61440 Schmitten, 861m; 50.233780, 8.459419; leg. F. Dal Grande & I. Schmitt October 2012; physodic acid chemotype	N.A.	SRS2339646
Pseudevernia furfuracea	metatranscriptome	thallus	Germany, Hesse, Taunus, Großer Feldberg, 61440 Schmitten, 861 m; 50.233780, 8.459419; leg. F. Dal Grande & I. Schmitt, July 2013; physodic acid chemotype	FR-0265084	SRS2339645
Pseudevernia furfuracea	metatranscriptome	thallus	Spain, Guadalajara, 19223 Majaelra yo, 1359 m; 41.141758, -3.306956; leg. A. Crespo, F. Dal Grande & P. K. Divakar, June 2012; olivetoric acid chemotype	FR-0265085	SRS2339647

Table 1. Specimen information and accession numbers of genomes and transcriptomes generated in this study (FR: Herbarium Senckenbergianum, Senckenberg Forschungsinstitut und Naturmuseum, Frankfurt/M, Germany; AKPM: Akita Prefectural Museum, Japan).

Reference genome assemblies from culture. Reads from pure fungal cultures were adapter- and quality trimmed as follows: for the paired-end libraries we used Trimmomatic[36] v0.33 with 2 seed mismatches, a palindrome clip threshold of 30 and a simple clip threshold of 10 for adaptor removal, a length cut-off of 60 for 100 bp reads and of 127 for 150 bp reads, removing leading and trailing low quality bases below a quality of 3 and a quality cut-off of an average 20 in a 5-base-wide sliding window. For the mate-pair libraries we used NxTrim[37] v0.3.2 instead of Trimmomatic. Additionally, Sickle[38] v1.33 was used with a 20-Phred quality threshold and length filter of 60 or 127 and *ecc.sh* in BBMap[39] v35.14 was used with default settings for error correction. The trimmed and filtered short-insert reads of *P. furfuracea* were overlapped with PEAR[40] v0.9.6. After some preliminary tests, we chose the best performing assembler for each species. For *E. prunastri* we used SPAdes[41] v3.5.0 with the recommended settings *careful* and *k*-mer length *21,33,55,77*. For *P. furfuracea* we used omega[42] v1.0.2 and a minimum overlap length of 60. We scaffolded contigs with SSPACE[43] v3.0 and used GapFiller[44] v1.10 to close remaining gaps. The resulting scaffolds were assigned taxonomically to Ascomycota with MetaWatt[45] v.3.5.2 to filter out potential contaminants. Assembly statistics were accessed with Assemblathon[46] and the genome completeness was estimated based on evolutionarily-informed expectations of gene content with BUSCO v.2.0 (Benchmarking Universal Single-Copy Orthologs)[47] and a lineage-specific set of Ascomycota single-copy orthologs from OrthoDB[48] v.9.

Reference gene sets from culture. We performed *de novo* gene prediction and annotation on the assemblies based on pure fungal cultures using MAKER[49] v2.31.8 in an iterative fashion following the recommendation and protocols of Campbell *et al.*[50] and incorporating the metatranscriptome data quality filtered with Trimmomatic v.0.36 and aligned with bowtie[51] v2.1.0. For the first round of MAKER we used Hidden Markov Models (HMMs) gained from GeneMark-ES[52] v4.33 and SNAP[53] with hints from CEGMA[54] v2.4 (performed on iPlant[55]) and included RNA evidence through a TopHat[56] v2.0.11 GFF file. Then we converted the first-round results to new SNAP and Augustus[57] v3.0.2 HMMs and ran MAKER again. Additionally, we rescued rejected gene models (MAKER *standard* instead of *default build*) including all gene models that were supported by RNA evidence and all *ab initio* gene models encoding a protein family (Pfam) domain detected by InterProScan[58] v.5.23–62.0 and that did not overlap with RNA evidence. For *P. furfuracea* we used the RNA evidence originating from the olivetoric acid chemotype to match the chemotype of the fungal culture. Gene set completeness was estimated as genome completeness with BUSCO (see above).

Metagenomic assemblies. For metagenomic reads we tested different *de novo* assemblers for whole-genome shotgun sequence data to evaluate their performance and obtain the best possible assembly[29,46]. We used a range of different *de novo* assemblers relying on the detection of overlapping reads (overlap layout graph assemblers) as well as those utilizing de Bruijn graphs and included general-purpose assemblers as well as specialized metagenome assemblers. We used the following assemblers: MIRA[59] v4.0.2, omega, SPAdes v3.8.1, metaSPAdes[9] v3.8.1, metaVelvet[60] v1.2.02 and IDBA-UD[10] v1.2.0. MIRA was run with the default flags *genome*, *denovo*, *accurate* and an auto refining template size of minimum 151 to maximum 600 on reads overlapped with

PEAR. Adaptor removal and quality trimming was not performed for MIRA following the developer's recommendation. For the other assemblers, we adapter-trimmed and quality-filtered the raw reads as described above, but using Trimmomatic v.0.36 with a length cut-off of 150 and Sickle with a length filter of 127. Trimmed and filtered reads were overlapped with PEAR and used as input to the assemblers described hereafter. SPAdes was run with the recommended settings *careful* and *k*-mer length *21,33,55,77,99,127*. MetaSPAdes was run with the flag *meta* for metagenomic samples and the same *k*-mer length settings as in SPAdes. MetaVelvet was used with Velvet[61] v1.2.10 assembling *k*-mer sizes from 51 to 231 with a step size of 20 using an estimated mean insert size of 273 for *E. prunastri* and 267 for *P. furfuracea* as calculated with bowtie v2.2.5 and a custom Python script. No consistent paired-end connection was chosen as recommended for metagenomic samples containing very dissimilar species and the expected coverage was set to *auto* initially and set manually in a re-run after inspection of *k*-mer coverage histograms plotted with the package plotrix[62] v3.6–4 in R[63] v3.3.2. The optimal metaVelvet *k*-mer sizes of 191 for *E. prunastri* and of 91 for *P. furfuracea* were chosen according to the VelvetOptimiser[64] v2.2.5 manual and Greshake *et al.*[29] by multiplying the N50 by the number of long contigs (>1 kbp). We tested omega with minimum overlap lengths between 100 and 200 and applied the same optimisation criteria as for the metaVelvet assemblies resulting in an overlap length of 140 for *E. prunastri* and 150 for *P. furfuracea*. IDBA-UD was chosen instead of meta-IDBA[65] as it generally performs better according to the authors and was run with *k*-mer sizes from 51 to 231 with a step size of 20. All resulting assemblies of all assemblers were filtered for a minimum length of 400 bp using a custom Perl script.

Taxonomic assignment. We compared two different approaches to extract fungal contigs from all metagenomic assemblies. For the first approach we ran DIAMOND[66] v0.8.34.96 BLASTx with the *more-sensitive* mode for longer sequences and a default e-value cut-off of 0.001 against the NCBI Genbank *nr* protein database[67] (downloaded in January 2017). We parsed the results with MEGAN[68] v.6.7.7 with *max expected* set to 1E-10 and using the weighted lowest common ancestor (LCA) algorithm which improves the specificity of taxonomic assignment compared to the naive LCA algorithm[69]. For all assemblies, we exported all contigs assigned to Ascomycota to represent the expected mycobiont[15]. An assignment to lower taxonomic rank was not possible due to the lack of closely related genomes in the reference databases. All downstream analyses refer to the extracted Ascomycota subsets.

For the second approach, we used MetaWatt. While MEGAN classifies reads based on sequence similarity by finding the LCA in the NCBI taxonomy, MetaWatt makes use of multivariate statistics of tetranucleotide frequencies and differential coverage based binning of metagenomic contigs. MetaWatt also performs taxonomic profiling of bins with DIAMOND BLASTx against a database that we customized to include non-redundant genomes of 532 Eukaryota, 1936 Bacteria, and 132 Archaea at the genus rank and 619 viruses at family rank (generated in August 2017). We calculated read coverage by aligning the quality trimmed reads with bowtie v2.2.5, converting files with samtools v.1.1[70] and running Qualimap v.2.2.1[71]. We disabled coverage based binning since we had only one read set as recommended in the manual. To identify tetranucleotide bins that belong to the respective mycobiont we used the following approach: we selected bins with an Ascomycota profile that had at least 50% of the fragments classified as Ascomycota and no other taxa represented in their taxonomic profile. We then merged all bins that met these criteria and manually unbinned contigs that were not classified as Ascomycota or 'Unknown' (see Supplementary Table S1).

Assembler evaluation. To find the best fungal assembly from metagenomic reads, we assessed quality and completeness of the assemblies extracted with MEGAN and MetaWatt using as reference genomes and gene sets the assemblies based on pure fungal cultures in QUAST[72] v.4.1. QUAST evaluates and compares assemblies based on alignment of contigs to references. We used a lower contig length threshold of 400 and the settings *scaffolds*, *eukaryote* and *fragmented*. We considered several assembly statistics and metrics from QUAST, e.g., number of contigs, total length and N50 (see full QUAST reports in the Supplementary File S6 or S7), number of misassemblies and the fraction of reference genome and genes covered compared to the reference genome. Additionally, we confirmed the QUAST evaluation with genome assembly and gene set completeness with BUSCO as described above. BUSCO results were visualized with the package ggplot2[73] v.2.2.1 in R.

Gene prediction and comparisons of genes sets. For comparing the gene sets of the reconstructed genomes from metagenomic reads with the gene sets of reference fungal cultures, we applied the following three steps to the best reconstructed genomes extracted with MEGAN and MetaWatt and for both species, respectively. First, we performed a *de novo* gene prediction and annotation on the fungal contig subsets using MAKER as described above, but using the RNA evidence originating from the physodic acid chemotype for P. furfuracea to match the chemotype of the mycobiont reconstructed from metagenome.

Second, we used Reciprocal Best Blast Hits (RBH) to find orthologous pairs between the gene sets, as a simple and fast method for comparing different gene sets resulting from different assemblies of the same species[74]. We ran BLASTp[75] v2.2.30+ using final Smith-Waterman alignment and soft filtering (BLAST flags *use_sw_tback*, *soft_masking true*, *seq yes* and *evalue 1e-6*) for better detecting orthologs as RBH[74,76]. We filtered the BLAST results for a minimum identity of 70% over the alignment length (*pident*) and a minimum query coverage of 50% (*qcovhsp*)[77], sorted for highest bit-score and lowest e-value and removed multiple identical top hits after manual inspection as not informative in the context of RBH. We used eulerAPE[78] v.3.0.0 to draw area-proportional Venn diagrams to visualize the overlap among gene sets.

Third, we tested if metagenome-based assemblies are reliable to study fungal gene space when a fungal culture is not available. For this purpose, we chose to analyse the gene families of secondary metabolism as they represent typical targets for focused genome mining of lichen-forming fungi[19,79]. We thus identified and annotated

secondary metabolite biosynthesis gene clusters with the fungal version of antiSMASH[80] v.4.0.2, including polyketide synthases, non-ribosomal peptide synthetases, and terpene synthases. We used an annotated nucleotide file in EMBL format as input generated from the genome FASTA file and the GFF file from MAKER with scripts provided by M. H. Medema. We compared the outputs in terms of total genes and RBHs.

Taxonomic composition of the metagenomic read sets. We investigated the taxonomic composition of the metagenomic read sets to estimate the abundance of the target mycobionts. For this, we taxonomically classified the trimmed and corrected metagenomic reads using Kraken[81] v.0.10.5-beta. Kraken utilises exact alignments of overlapping read *k*-mers and is one of the best tools in terms of precision and accuracy[82]. Here we took advantage of the fact that we could incorporate the reference genomes of our fungal species into a custom-build database. It was therefore possible to estimate with precision the amount of reads of the lichen-forming fungus and compare it to that of other fungi in each metagenome. The Kraken database we built thus contained our reference genomes of *E. prunastri* and *P. furfuracea*, and the entire fungal RefSeq[83] (release 79). We further included five lichen genomes (*Cladonia macilenta*, *Cladonia metacorallifera*, *Endocarpon pusillum*, *Gyalolechia flavorubescens*, *Umbilicaria muehlenbergii*) and the basidiomycete *Cystobasidium pallidum*[15] from NCBI GenBank. We hard-masked the database genomes for low-complexity regions with dustmasker v.1.0.0 (part of BLAST v.2.2.30) as recommended in the manual. We then built and classified the database for a read length of 250 bp.

To identify potential sources of contamination, we further taxonomically classified all metagenomic reads that were not classified as reference mycobiont by the above-mentioned Kraken approach. For this, we ran DIAMOND BLASTx with default settings against the NCBI Genbank *nr* protein database and parsed the results in MEGAN with *max expected* set to 0.001 and using the weighted LCA algorithm. Additionally, we extracted the reads that were assigned to non-target Ascomycota in MEGAN, assembled these as described above with SPAdes and searched for the presence of biosynthetic genes in the resulting scaffolds with antiSMASH.

Results

Sequencing results. For *E. prunastri* we obtained a total of 104,138,074 paired-end and mate-pair reads from culture, 21,622,755 paired-end reads from metagenome and 32,706,203 paired-end reads from metatranscriptome. For *P. furfuracea* we obtained a total of 70,695,549 paired-end and mate-pair reads from culture, 21,031,517 paired-end reads from metagenome, 17,357,422 paired-end reads from the olivetoric acid chemotype metatranscriptome and 18,321,601 paired-end reads from the physodic acid chemotype metatranscriptome.

Reference genomes and gene sets. We first obtained references genomes for *E. prunastri* and *P. furfuracea* from fungal cultures. After quality filtering and trimming, we used 85.7% *E. prunastri* reads and *P. furfuracea* 79.3% reads for genome assembly. The genome of *E. prunastri* was assembled into 277 scaffolds with total length of 40 Mbp, N50 of 264,454 bp, an average coverage of ~410x, and an estimated genome completeness of 95.9% according to BUSCO. The genome of *P. furfuracea* was assembled into 46 scaffolds with a total length of 38 Mbp, N50 of 1,178,799 bp, average coverage of ~350x, and estimated 94.7% completeness. After quality trimming, we obtained 87.9% paired-end RNA reads of *E. prunastri* of which 55.4% aligned to the genome to give hints for gene prediction. For *P. furfuracea* 90.8% RNA reads survived quality filtering and 57.1% of these mapped against the genome. We predicted 10,992 genes for *E. prunastri* and 8,842 genes for *P. furfuracea* with an estimated gene set completeness of 92.1% and 91.8%, respectively. Both reference genomes and gene sets based on the pure culture strains of *E. prunastri* and *P. furfuracea* are summarized in Table 2.

Evaluation of metagenomic assemblies. After quality filtering and trimming, 29,573,575 (68.4%) *E. prunastri* metagenomic reads and 35,116,468 (83.4%) *P. furfuracea* metagenomic reads were used with six assemblers. We evaluated assembler performance based on the extracted fungal subset with MEGAN and MetaWatt based on overall genome statistics, overlap to reference, and fungal genome completeness. We observed highly different assemblies depending on the assembler used, while the choice of taxonomic assignment method did not yield significant differences (Fig. 1). Comparisons of N-statistics and the covered fraction of the reference genome are shown in Fig. 1. Full QUAST reports are provided in the Supplementary Files S6 and S7. SPAdes assemblies had the best assembly statistics and the highest overlap to the reference (86–90% for *E. prunastri* and 80–87% for *P. furfuracea*). MetaSPAdes was second best followed by IDBA-UD. We compared the fungal genome completeness based on BUSCO of all unassigned, MEGAN-assigned and MetaWatt-assigned assemblies in Fig. 2 (detailed values can be found in Supplementary Table S2). The assessment of completeness was based on a lineage-specific set for Ascomycota and therefore provides the possibility to access 'potential' completeness in the taxonomically unassigned assemblies. We observed that the fungal completeness in unassigned assemblies (95.5% for *E. prunastri*, 94.0% for *P. furfuracea*) was comparable to the MEGAN-assigned assemblies, while MetaWatt-assigned assemblies were slightly less complete (see Fig. 2 and Supplementary Table S2). The SPAdes assembly had the highest fungal completeness in MEGAN-assigned assemblies and thus corroborated the QUAST results. For MetaWatt-assigned assemblies, MetaSPAdes had a slightly higher fungal completeness compared to SPAdes, but, based on QUAST, SPAdes had overall better genome statistics and higher overlap to the reference genome. MEGAN generally assigned a higher number of contigs to Ascomycota compared to MetaWatt, although genome completeness did not deviate greatly between the two methods. We found a large core overlap of contigs assigned to Ascomycota by both taxonomic assignment methods (Supplementary Table S3). Overall, genome statistics, overlap to reference and fungal genome completeness showed the same trends in both examined species. For downstream analyses, we selected SPAdes assemblies assigned in MEGAN and MetaWatt for both species as best-possible reconstructed mycobiont genomes from metagenomic sets. These assemblies showed similar completeness as the reference genomes, but were, as expected, more fragmented as indicated by the higher number

	Evernia prunastri	*Pseudevernia furfuracea*
Number of scaffolds	277	46
Total size	40 Mb	38 Mb
N50	264,454 bp	1,178,799 bp
Average coverage	~410x	~350x
Number of genes	10,992	8,842
Genome completeness	C:95.9% [S:95.1%, D:0.8%], F:2.7%, M:1.4%	C:94.7% [S:94.6%, D:0.1%], F:3.5%, M:1.8%
Gene set completeness	C:92.1% [S:91.3%, D:0.8%], F:5.1%, M:2.8%	C:91.8% [S:91.8%, D:0.0%], F:5.0%, M:3.2%

Table 2. Reference genome assemblies and gene sets from pure culture strains. Completeness is compared against 1,315 orthologous BUSCO marker genes for Ascomycota (C:complete [S:single-copy, D:duplicated], F:fragmented, M:missing).

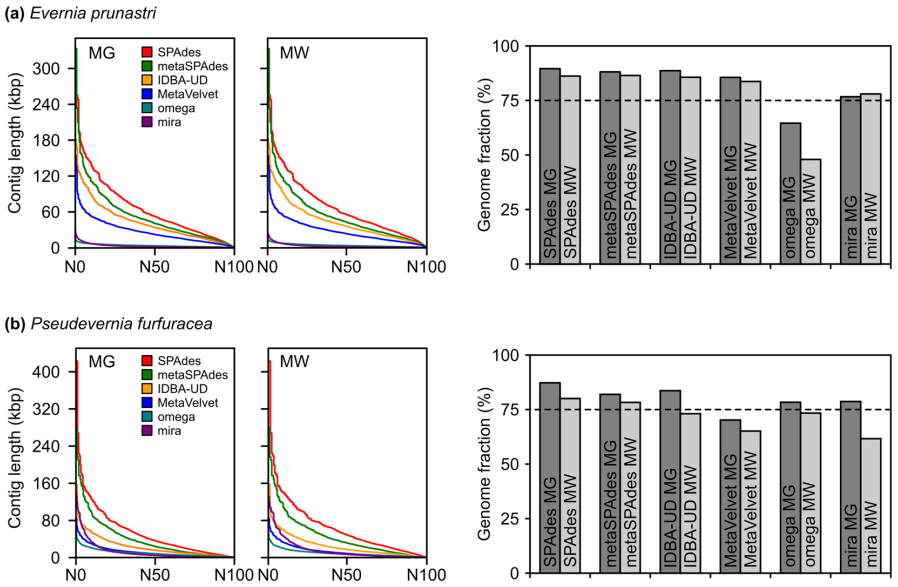


Figure 1. Comparison of six assembler s. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW). N-Statistics are presented on the left side and the fractions of the fungal culture reference genome that is covered by the metagenome-based assemblies are presented on the right side. This analysis is based on QUAST. Full QUAST reports are provided in the Supplementary Files S6 and S7.

of scaffolds and the lower N50 (Table 3). SPAdes assemblies of *E. prunastri* had an average coverage of ~110x (MEGAN) and ~105x (MetaWatt), while SPAdes assemblies of *P. furfuracea* had an average coverage of ~135x (MEGAN) and ~160x (MetaWatt).

Comparison of gene sets. After quality trimming, we obtained 28,755,102 (87.9%) paired-end RNA reads of *E. prunastri* of which 53.9% aligned to the MEGAN- and 51.3% aligned to the MetaWatt-assigned metagenome to give hints for gene prediction. For *P. furfuracea* 16,649,331 (90.9%) paired-end RNA reads remained after quality filtering and 58.6% of these mapped against the MEGAN-assigned metagenome assembly, and 53.2% mapped against the MetaWatt-assigned assembly. We predicted 11,098 genes from the MEGAN-assigned assembly and 10,713 genes from the MetaWatt-assigned assembly for *E. prunastri*. For *P. furfuracea* we predicted 10,028 genes from the MEGAN-assigned assembly and 8,962 genes from the MetaWatt-assigned assembly. These four gene sets had completeness between 89.2% and 91.7% (Table 3) which is comparable to the reference gene sets.

We identified orthologous pairs (RBH) among the different gene sets of each species. The reference genes overlapped to a large extent with the genes from the MEGAN- and the MetaWatt-assigned metagenome assemblies with a core overlap between the three of 87% for *E. prunastri* and 83% for *P. furfuracea* (Fig. 3). The MEGAN set covered 90.2% and MetaWatt covered 88.9% of the pure fungal culture genes in *E. prunastri*. In *P.*

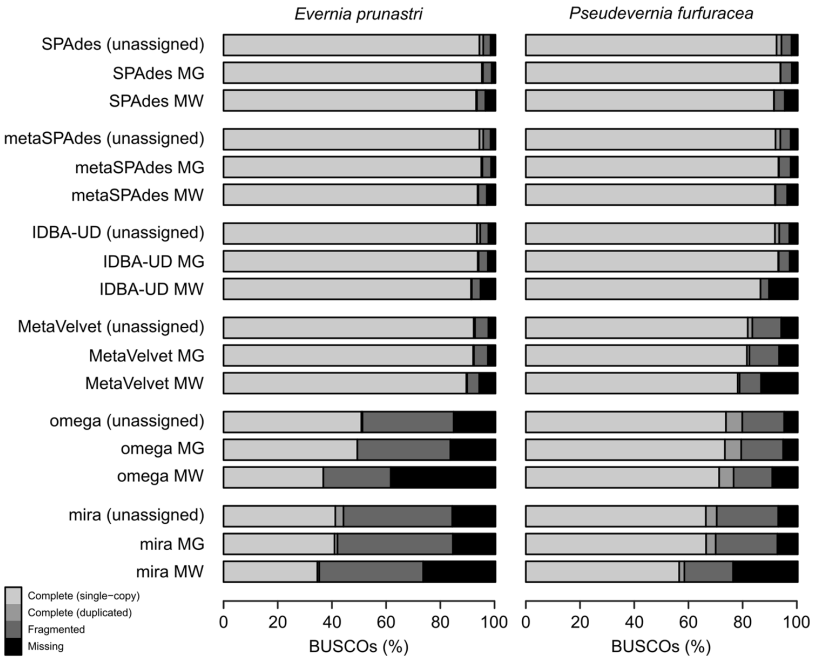


Figure 2. Genome completeness for genome assemblies from metagenomic lichen thalli using different assemblers. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW). Percentage completeness is compared against 1,315 orthologous BUSCO marker genes for Ascomycota. Full BUSCO reports are provided in Supplementary Table S2.

	Evernia prunastri		*Pseudevernia furfuracea*	
	metagenome MG	**metagenome MW**	**metagenome MG**	**metagenome MW**
Number of scaffolds	1,775	1,624	3,558	1,829
Total size	39 Mb	37 Mb	43 Mb	33 Mb
N50	53,038 bp	54,988 bp	36,386 bp	48,187 bp
Average coverage	~110x	~105x	~135x	~160x
Number of genes	11,098	10,713	10,028	8,962
Genome completeness	C: 95.4%	C: 93.2%	C: 93.7%	C: 91.3%
Gene set completeness	C: 91.7%	C: 89.2%	C: 91.3%	C: 89.4%

Table 3. Fungal genome assemblies and gene sets from metagenomic lichen thalli assembled with SPAdes. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW). Completeness is compared against 1,315 orthologous BUSCO marker genes for Ascomycota (C: complete BUSCO genes found).

furfuracea the MEGAN set covered 87.8% and MetaWatt covered 85.6% of the pure fungal culture genes. The metagenomic-based gene sets overlapped with 10,269 genes in *E. prunastri* and 8,496 genes in *P. furfuracea*.

We identified a high number of reducing and non-reducing polyketide synthases, non-ribosomal peptide synthetases and terpene synthases with antiSMASH in both species. In total, we found 50 biosynthetic genes in the *E. prunastri* culture-based reference genome and 51 and 49 biosynthetic genes in the MEGAN- and MetaWatt-assigned metagenome assemblies. In *P. furfuracea* we found 31 biosynthetic genes in the reference and 31 and 27 biosynthetic genes in the MEGAN- and MetaWatt-assigned assemblies, respectively. We observed a high overlap of genes between biosynthetic genes found in culture and in metagenomic samples. In *E. prunastri* 82% of the biosynthetic genes in the culture (41 of 50) were also present in the metagenome, while we found 71% (22 of 31) of the biosynthetic genes in the metagenome of *P. furfuracea* (Table 4).

Taxonomic composition of metagenomic reads. The taxonomic classification and abundance estimation of quality-filtered metagenomic reads is shown in Fig. 4. Kraken estimated 13,850,095 (73.4%)

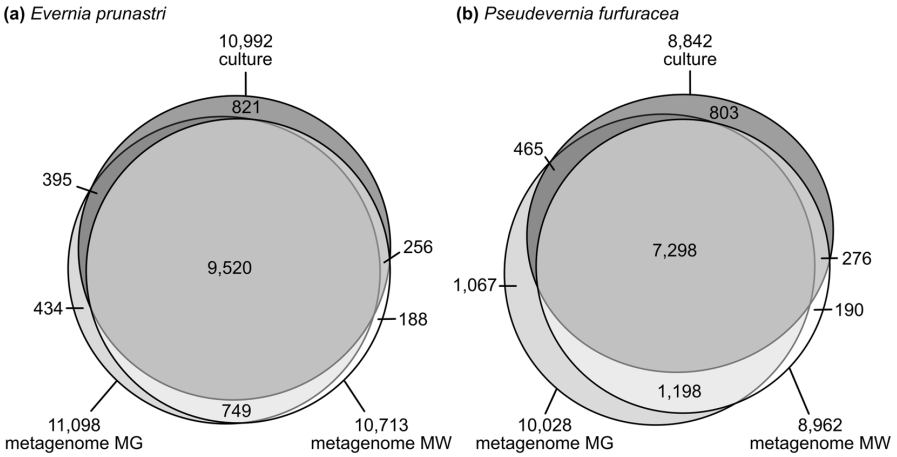


Figure 3. Orthologous gene sets in assemblies from fungal culture and from natural lichen sample. Analyses are based on Reciprocal Best Hits. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW).

	Evernia prunastri				*Pseudevernia furfuracea*			
	culture	**metagenome MG**	**metagenome MW**	**common**	**culture**	**metagenome MG**	**metagenome MW**	**common**
Reducing type I PKS	20	20	20	19	12	9	7	6
Non-reducing type I PKS	8	8	7	6	5	6	6	5
Type III PKS	2	2	2	2	2	2	2	2
Hybrid PKS-NRPS	4	3	3	3	3	3	3	2
NRPS	4	4	5	3	4	5	4	3
Terpene synthases	12	14	12	8	5	6	5	4
Total	50	51	49	41	31	31	27	22

Table 4. Biosynthetic genes identified in cultures, the MEGAN-assigned metagenomes (MG) and the MetaWatt-assigned metagenomes (MW). The metagenome derived assemblies were taxonomically assigned to Ascomycota. PKS (Polyketide synthases), NRPS (Non-ribosomal peptide synthetases). Type III PKSs refer to chalcone synthases.

metagenomic reads to belong to *E. prunastri* and 14,178,239 (72.2%) to *P. furfuracea*. A further classification of the 26–27% of reads that were not classified as reference mycobionts with BLAST/MEGAN against the entire GenBank *nr* protein database is also shown in Fig. 4. In the metagenome of *E. prunastri* we found 10.7% Bacteria, 0.7% Fungi (0.5% Ascomycota, 0.2% Basidiomycota) and 0.6% Viridiplantae. In *P. furfuracea* we found 1.9% Bacteria, 1% Fungi (0.9% Ascomycota, 0.1% Basidiomycota) and 2.1% Viridiplantae. A large proportion of the reads (14.2% for *E. prunastri* and 22.4% for *P. furfuracea*) were of unknown origin, i.e. without a BLAST hit or not assignable in MEGAN. Detailed read numbers of the classification with MEGAN are given in the Supplementary Table S4.

The search for biosynthetic genes in the assembled proportion of non-target Ascomycota resulted in one non-ribosomal peptide synthetase in the *E. prunastri* metagenome and one hybrid polyketide non-ribosomal peptide synthetase in the metagenome of *P. furfuracea*.

Data availability. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accessions NKYR00000000 (*Evernia prunastri* mycobiont culture) and NKYQ00000000 (*Pseudevernia furfuracea* mycobiont culture). The metagenomic sequence data and metatranscriptomes are available under the accession number SRP111200. The metagenomic assemblies are available at figshare (https://doi.org/10.6084/m9.figshare.5531692).

Discussion

We evaluated the applicability of the metagenome skimming approach to assemble genomes of lichen-forming fungi by comparing genomes assembled from mixed DNA samples (natural thalli) to genomes assembled from DNA from pure fungal cultures. The metagenome-derived fungal assemblies are comparable to the reference genomes from pure culture in terms of total genome size and genome completeness. While the

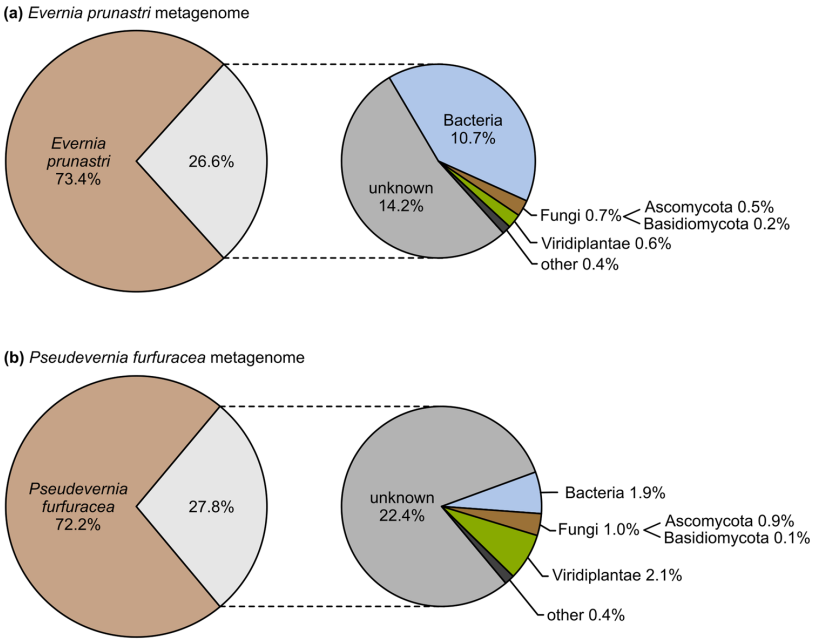


Figure 4. Taxonomic classification and abundance estimation of the quality filtered metagenomic reads from lichen thalli. For the pie chart on the left the reads were assigned using Kraken against a custom database of fungi to assess how many reads belong to the reference lichen-forming fungus. For the pie chart on the right, we classified all reads that did not belong to the lichen-forming fungus using BLAST/MEGAN with the NCBI *nr* protein database. The 'unknown' proportion refers to reads without a BLAST hit or reads that could not be assigned in MEGAN. Detailed read numbers can be found in Supplementary Table S4.

metagenome-derived genomes from a single shotgun sequencing library are naturally more fragmented than the respective references that were sequenced at a deeper coverage and with different sequencing libraries, they still cover most of the reference gene space (estimated 86–90%).

Comparison of metagenomic assembly strategies. We observed extensive differences in assembly quality among assemblers. To some extent this might be due to the presence of highly uneven read coverages. Different species in a metagenomic sample have different abundances resulting in a highly non-uniform read coverage across different genomes. Furthermore, the coverage of most species from a metagenomic sample is much lower than in a typical sequencing project of a cultivated sample[7,9]. Lichen thalli represent multi-species communities of fungal, algal and bacterial species[14] that were shown to have uneven coverages and therefore pose challenges for assemblers[29]. The best performing assemblers, SPAdes, MetaSPAdes and IDBA-UD, were to some extent especially designed for sequencing data with highly uneven coverage[9,10,41], while in contrast omega and MetaVelvet are surprisingly negatively affected by such data[29]. The latter two employ coverage information to distinguish between species and therefore might have difficulties in assembling the fungal genome from uneven coverage short-read sequencing data[84].

Our study was designed to compare metagenomic assemblies against pure fungal reference genomes. Our results suggest that a reference is not necessarily needed to choose the best assembler, as assembly statistics and fungal genome completeness alone would have led to the selection of the same assembler. SPAdes, or its metagenomic variant MetaSPAdes, seem to be among the most reliable (and fastest) tools for assembling metagenomic reads in the presence of highly uneven read coverage.

Comparison of metagenomic binning approaches. The choice of taxonomic binning method (MEGAN and MetaWatt) made no large difference, but MEGAN performed slightly better in terms of mycobiont completeness. On the one hand, the BLAST/MEGAN approach is commonly used[29,85,86], and constant development and added functionality (e.g., functional profiling and direct comparison of several samples) have improved MEGAN over time[68,69,87]. MEGAN relies on an initial BLAST run against the NCBI *nr* database which requires considerable disk space for the database and computational power for the DIAMOND search. On the other hand, MetaWatt requires considerably less time for a similarly good taxonomic assignment as it can already be efficiently used with a smaller database of genomes by deploying multivariate statistics of tetranucleotide frequencies and differential coverage based binning rather than sequence similarity.

Comparison of gene space. We were able to cover up to 88–90% of the respective reference gene spaces. Furthermore, we showed that metagenome-derived assemblies reliably recovered almost all members of the diverse gene families involved in secondary metabolism. We showed a high diversity of secondary metabolite genes in *E. prunastri* and *P. furfuracea* as expected from the high number of substances that have been reported from extracts of these species[31,32]. The high overlap of these biosynthetic genes in the metagenomes compared to their reference suggests a great potential of the metagenome skimming approach for natural product discovery in lichens[88]. Beside mining the genomes for secondary metabolites that produce interesting lichen compounds, the potential of whole genome mining can be extended to other target gene families, e.g. mating-type genes, symbiosis-related proteins or secreted effector proteins[89].

A few secondary metabolism-related genes were exclusively found in the metagenomes. These genes may result either from the annotation of more fragmented assemblies or, as shown in two cases, from contamination, e.g. from other non-lichen-forming ascomycetes (i.e., lichenicolous fungi). By using the fungal genomes obtained from aposymbiotic cultures, we were able to access the taxonomic composition of reads used for the metagenomic assemblies. Interestingly, nearly three-quarters of our metagenomic reads represent the lichen-forming fungus while other ascomycetes constituted less than 0.9% of the total reads. This suggests that only a minor fraction of the additional fungal gene models found in the metagenomes may represent contamination from other fungi that could not be filtered out during taxonomic assignment due to incompleteness of the reference database.

One of the essential steps in reconstructing genomes from metagenomic samples is the taxonomic assignment of sequences[90]. Sensitivity and precision of assignment methods depend on the availability and quality of reference databases and taxonomies[91,92]. Taxonomic binning of metagenomic sequences will improve in the future when more genomes become available in public databases and the development of sequence composition-based methods in comparison to similarity-based methods will progress further. Until then, without good reference databases it is not possible to entirely exclude contamination in metagenomic assemblies or target specific genomes by filtering out unwanted sequences.

Additional applications and conclusions. Our results show that metagenome skimming constitutes a comprehensive genome-mining tool for lichens, and potentially for other microbial symbioses. Fungal genomes reconstructed from metagenomic read sets can be used in comparative phylogenomics, an approach to link genomic features to traits in a phylogenetic context[93]. A recent study showed that phylogenomic datasets can be useful to resolve evolutionary relationships among cryptic lichen-forming fungal lineages[94]. In this context, metagenome skimming can drastically extend the traditional set of DNA barcodes on which to infer phylogenetic relationships, while circumventing the necessity for a reference genome from culture[95,96]. This is especially important for filling gaps in the tree of life as some species rich-clades in the Ascomycota consist entirely of lichenized fungi (Lecanoromycetes, Lichinomycetes, and a large subset of Eurotiomycetes). Additionally, it will help to better understand the impact of lichenization on the evolution of fungal genomes.

Another significant potential of metagenome skimming lies in the development of high-resolution population genetic markers. A sensitive fingerprinting technique is especially important in highly clonal organism such as lichens[97]. Here, microsatellites can be used to investigate microevolutionary processes at the level of populations to study association patterns or reconstruct gene flow and symbiont transmission[86,98,99]. In recent years, an increasing number of studies have used genome skimming short read libraries for marker isolation, in particular fungus-specific microsatellites (SSRs). Metagenome skimming is in fact cheaper and more informative than the library microsatellite enrichment approach, as it retrieves all kinds of repeats (microsatellites, minisatellites, and potentially also transposable elements). Furthermore, here we have shown that the metagenome skimming approach covers most of the gene space of the fungal symbiont.

In conclusion, our findings suggest that metagenome skimming is a viable tool to reconstruct nearly complete fungal genomes from uncultured lichen thalli. Additionally, algal and bacterial contigs can be accessed via taxonomic binning and functional annotation to explore biodiversity and natural compound diversity[23,88]. Metagenome skimming circumvents the time-consuming step of cultivation and can be applied to unculturable organisms, symbiotic consortia, and other complex communities. It therefore bears a great potential to be applied in several research fields that are rapidly changing our view on the ecology and evolution of symbiotic associations.

References

1. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
2. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
3. Juengst, E. & Huss, J. From metagenomics to the metagenome: Conceptual change and the rhetoric of translational genomic research. *Genomics, Soc. Policy* **5**, 1–19 (2009).
4. Abram, F. Systems-based approaches to unravel multi-species microbial community functioning. *Comput. Struct. Biotechnol. J.* **13**, 24–32 (2015).
5. Howe, A. C. *et al.* Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl. Acad. Sci.* **111**, 4904–4909 (2014).
6. Franzosa, Ea *et al.* Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* **13**, 360–372 (2015).
7. Nagarajan, N. & Pop, M. Sequence assembly demystified. *Nat. Rev. Genet.* **14**, 157–167 (2013).
8. Daniel, R. The metagenomics of soil. *Nat. Rev. Microbiol.* **3**, 470–478 (2005).
9. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
10. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
11. Grube, M. & Wedin, M. Lichenized Fungi and the Evolution of Symbiotic Organization. *Microbiol. Spectr.* **4**, 1–17 (2016).
12. Nash, T. H. *Lichen Biology*. (Cambridge University Press, 2008).
13. Cernava, T., Berg, G. & Grube, M. High Life Expectancy of Bacteria on Lichens. *Microb. Ecol.* **72**, 510–513 (2016).
14. Aschenbrenner, I. A., Cernava, T., Berg, G. & Grube, M. Understanding Microbial Multi-Species Symbioses. *Front. Microbiol.* **7**, 180 (2016).
15. Spribille, T. *et al.* Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**, 488–492 (2016).

16. Fernández-Mendoza, F., Fleischhacker, A., Kopun, T., Grube, M. & Muggia, L. ITS1 metabarcoding highlights low specificity of lichen mycobiomes at a local scale. *Mol. Ecol.* **38**, 42–49 (2017).
17. Dal Grande, F. *et al.* Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol.* https://doi.org/10.1111/nph.14770 (2017).
18. Huneck, S. & Yoshimura, I. *Identification of Lichen Substances*. (Springer-Verlag, 1996).
19. Huneck, S. The significance of lichens and their metabolites. *Naturwissenschaften* **86**, 559–570 (1999).
20. Boustie, J. & Grube, M. Lichens - a promising source of bioactive secondary metabolites. *Plant Genet. Resour.* **3**, 273–287 (2005).
21. Molnár, K. & Farkas, E. Current results on biological activities of lichen secondary metabolites: a review. *Zeitschrift für Naturforsch. C* **65C**, 157–173 (2010).
22. McDonald, T. R., Mueller, O., Dietrich, F. S. & Lutzoni, F. High-throughput genome sequencing of lichenizing fungi to assess gene loss in the ammonium transporter/ammonia permease gene family. *BMC Genomics* **14**, 225 (2013).
23. Kampa, A. *et al.* Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci.* **110**, E3129–E3137 (2013).
24. Erlacher, A. *et al.* Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of Lobaria pulmonaria L. *Front. Microbiol.* **6**, 1–9 (2015).
25. Lutsak, T., Fernández-Mendoza, F., Kirika, P., Wondafrash, M. & Printzen, C. Mycobiont-photobiont interactions of the lichen *Cetraria aculeata* in high alpine regions of East Africa and South America. *Symbiosis* **68**, 25–37 (2016).
26. Werth, S., Miao, V. P. W., Jónsson, Z. O. & Andrésson, Ó. S. High-throughput sequencing in studies of lichen population biology in *Recent Advances in Lichenology* (eds. Upreti, D. K. *et al.*) 61–94 (Springer, 2015).
27. Crittenden, P. D., David, J. C., Hawksworth, D. L. & Campbell, F. S. Attempted isolation and success in the culturing of a broad spectrum of lichen-forming and lichenicolous fungi. *New Phytol.* **130**, 267–297 (1995).
28. McDonald, T. R., Gaya, E. & Lutzoni, F. Twenty-five cultures of lichenizing fungi available for experimental studies on symbiotic systems. *Symbiosis* **59**, 165–171 (2013).
29. Greshake, B. *et al.* Potential and pitfalls of eukaryotic metagenome skimming: A test case for lichens. *Mol. Ecol. Resour.* **16**, 511–523 (2016).
30. McDonald, T. R. Genomic insights into the lichen symbiosis: *Cladonia grayi* as a model lichen. Ph.D. thesis, Duke University (2011).
31. Joulain, D. & Tabacchi, R. Lichen extracts as raw materials in perfumery. Part 1: oakmoss. *Flavour Fragr. J.* **24**, 49–61 (2009).
32. Joulain, D. & Tabacchi, R. Lichen extracts as raw materials in perfumery. Part 2: treemoss. *Flavour Fragr. J.* **24**, 105–116 (2009).
33. Beck, A. & Koop, H.-U. Analysis of the photobiont population in lichens using a single-cell manipulator. *Symbiosis* **31**, 57–67 (2001).
34. Cubero, O. F. & Crespo, A. Isolation of nucleic acids from lichens in *Protocols in Lichenology* (eds. Kranner, I. C. *et al.*) 381–391 (Springer, 2002).
35. Rubio-Piña, J. A. & Zapata-Pérez, O. Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants. *Electron. J. Biotechnol.* **14**, 1–8 (2011).
36. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
37. O'Connell, J. *et al.* NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics* **31**, 2035–2037 (2015).
38. Joshi, N. A. & Fass, J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33) [Software]. *Available at* https://github.com/najoshi/sickle (2011).
39. Bushnell, B. BBM (version 35.14) [Software]. *Available at https://sourceforge.net/projects/bbmap* (2015).
40. Zhang, J., Kobert, K., Flouri, T. X. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
41. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
42. Haider, B. *et al.* Omega: an Overlap-graph *de novo* Assembler for Metagenomics. *Bioinformatics* **30**, 2717–2722 (2014).
43. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
44. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **13**, R56 (2012).
45. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3**, 1–11 (2012).
46. Bradnam, K. R. *et al.* Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10 (2013).
47. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
48. Zdobnov, E. M. *et al.* OrthoDBv9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).
49. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
50. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **48**, 4.11.1-4.11.39 (2014).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
52. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).
53. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
54. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
55. Goff, S. A. *et al.* The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Front. Plant Sci.* **2**, 34 (2011).
56. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
57. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
58. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
59. Chevreux, B., Wetter, T. & Suhai, S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Proc. Ger. Conf. Bioinforma.* **99**, 45–56 (1999).
60. Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res.* **40**, e155 (2012).
61. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
62. Lemon, J. Plotrix: a package in the red light district of R. *R-News* **6**, 8–12 (2006).
63. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* Vienna, Autria. https://www.R-project.org/ (2015).
64. Gladman, S. & Seemann, T. VelvetOptimiser (version 2.25) [Software]. *Available at http://www.vicbioinformatics.com/software.velvetoptimiser.shtml* (2012).
65. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. Meta-IDBA: A *de Novo* assembler for metagenomic data. *Bioinformatics* **27**, 94–101 (2011).
66. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
67. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **45**, D37–D42 (2017).

68. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLOS Comput. Biol.* **12**, 4–12 (2016).
69. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
70. Li, H. *et al.* The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
71. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2015).
72. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
73. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2009).
74. Ward, N. & Moreno-Hagelsieb, G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS One* **9**, 1–6 (2014).
75. Altschul, S. F. *et al.* Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
76. Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).
77. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
78. Micallef, L. & Rodgers, P. euler APE: Drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9**, (2014).
79. Abdel-Hameed, M., Bertrand, R. L., Piercey-Normore, M. D. & Sorensen, J. L. Putative identification of the usnic acid biosynthetic gene cluster by *de novo* whole-genome sequencing of a lichen-forming fungus. *Fungal Biol.* **120**, 306–316 (2016).
80. Blin, K. *et al.* antiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36-W41 (2017).
81. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
82. Lindgreen, S., Adair, K. L. & Gardner, P. P. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci. Rep.* **6**, 19233 (2016).
83. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, 501–504 (2007).
84. Pop, M. Genome assembly reborn: recent computational challenges. *Brief. Bioinform.* **10**, 354–366 (2009).
85. Grube, M. *et al.* Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics. *ISME J.* **9**, 412–424 (2015).
86. Lutsak, T. *et al.* Characterization of Microsatellite Loci in the Lichen-Forming Fungus *Cetraria aculeata* (Parmeliaceae, Ascomycota). *Appl. Plant Sci.* **4**, 1600047 (2016).
87. Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).
88. Adnani, N., Rajski, S. R. & Bugni, T. S. Symbiosis-inspired approaches to antibiotic discovery. *Nat. Prod. Rep.* **34**, 784–814 (2017).
89. Wang, Y.-Y. *et al.* Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genomics* **15**, 34 (2014).
90. Dröge, J. & McHardy, A. C. Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief. Bioinform.* **13**, 646–655 (2012).
91. Santamaria, M. *et al.* Reference databases for taxonomic assignment in metagenomics. *Brief. Bioinform.* **13**, 682–695 (2012).
92. Peabody, M. A., Van Rossum, T., Lo, R. & Brinkman, F. S. L. Evaluation of shotgun metagenomics sequence classification methods using *in silico* and *in vitro* simulated communities. *BMC Bioinformatics* **16**, 363 (2015).
93. Delaux, P.-M. Comparative phylogenomics of symbiotic associations. *New Phytol.* **205**, 1443–1447 (2016).
94. Leavitt, S. D. *et al.* Resolving evolutionary relationships in lichen-forming fungi using diverse phylogenomic datasets and analytical approaches. *Sci. Rep.* **6**, 22262 (2016).
95. Papadopoulou, A., Taberlet, P. & Zinger, L. Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Mol. Ecol.* **24**, 3515–3517 (2015).
96. Coissac, E., Hollingsworth, P. M., Lavergne, S. & Taberlet, P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.* **25**, 1423–1428 (2016).
97. Arnaud-Haond, S., Duarte, C. M., Alberto, F. & Serrão, E. A. Standardizing methods to address clonality in population studies. *Mol. Ecol.* **16**, 5115–5139 (2007).
98. Dal Grande, F., Widmer, I., Wagner, H. H. & Scheidegger, C. Vertical and horizontal photobiont transmission within populations of a lichen symbiosis. *Mol. Ecol.* **21**, 3159–3172 (2012).
99. Alors, D. *et al.* Panmixia and dispersal from the Mediterranean Basin to Macaronesian Islands of a macrolichen species. *Sci. Rep.* **7**, 40879 (2017).

Acknowledgements

This project was supported by the research funding program Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz (LOEWE) of Hesse's Ministry of Higher Education, Research, and the Arts with grants to BiK-F and IPF. Some analyses were performed on the FUCHS cluster of the Center for Scientific Computing (CSC) of the Goethe University in Frankfurt am Main. We thank Claus Weiland (Frankfurt) for support in scientific computing, and Andreas Beck (Munich), Jae-Seoun Hur (Suncheon) and Yoshikazu Yamamoto (Akita) for help in obtaining fungal cultures. We also thank Carson Holt (MAKER), Marnix H. Medema (antiSMASH), Felipe A. Simão (BUSCO) and Marc Strous (MetaWatt) for support and advice using their software.

Author Contributions

A.M. and F.D.G. designed the study and analysed the data. I.S. and F.D.G. collected the samples from the field. J.O. performed cultivation and DNA/RNA isolation. A.M. and F.D.G. wrote the manuscript with input from I.S. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-14576-6.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material to

Meiser, A., Otte, J., Schmitt, I., & Dal Grande, F. (2017). Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi. *Scientific Reports*, 7(1), 14881, doi:10.1038/s41598-017-14576-6.

Table S1. Identification of tetranucleotide bins that correspond to the lichen-forming fungus with MetaWatt.
Table S2. BUSCO genome completeness of all fungal genomes assembled from metagenomic source and assigned to Ascomycota.
Table S3. Number of contigs of assemblies based on metagenomic reads and taxonomic assignment of metagenomic reads to Ascomycota with MEGAN and MetaWatt.
Table S4. Taxonomic assignment of quality filtered metagenomic reads from lichen thalli that were not assigned as the reference lichen-forming fungus.
Table S5. Taxonomic assignment of scaffolds assembled with SPAdes from metagenomic reads from lichen thalli.
File S6. Full QUAST reports of *Evernia prunastri* (HTML file).[1]
File S7. Full QUAST reports of *Pseudevernia furfuracea* (HTML file).[1]

[1] This supplementary material can be found online accompanying the published paper at http://dx.doi.org/10.1038/s41598-017-14576-6.

Table S1. Identification of tetranucleotide bins that correspond to the lichen-forming fungus with MetaWatt. Bins that had an Ascomycota profile with at least 50 % of the fragments classified as Ascomycota and no other taxa represented in their taxonomic profile were selected. All bins that met these criteria were merged and contigs that were not classified as Ascomycota or 'Unknown' were unbinned manually. Statistics presented here refer to the combined bins for each assembly and species.

Species		**SPAdes**	**metaSPAdes**	**IDBA-UD**	**MetaVelvet**	**omega**	**mira**
Evernia prunastri	Combined bins	3	4	3	4	5	16
	Removed contigs	68	29	58	34	72	814
	All contigs	1,624	1,602	2,757	2,432	7,577	39,338
	Ascomycota contigs	995 (61 %)	1,262 (79 %)	1,417 (51 %)	2,017 (83 %)	4,789 (63 %)	15,744 (40 %)
	'Unknown' contigs	629	340	1,340	415	2,788	23,594
	Size	36.8 Mb	36.4 Mb	36.8 Mb	35.3 Mb	20.2 Mb	58.7 Mb
	N50	54,988	42,862	35,676	23,489	3,602	1,679
	GC	49.3 - 50.6 %	49.4 - 50.7 %	49.1 - 50.6 %	49.2 - 50.6 %	49.8 - 51.8 %	48.8 - 51.5 %
	tRNA's	24	26	29	22	7	78
	Coding density	54.3 %	54.6 %	53.9 %	54.5 %	60.4 %	55.2 %
	Coverage	106.9 x	106.8 x	105.4 x	105.8 x	105.6 x	49.2 x
	Profile completeness	52.4 %	53.7 %	51.8 %	54.1 %	34.5 %	52.5 %
	Profile copies	1.6 x	1.6 x	1.6 x	1.6 x	1.4 x	2.1 x
	Abundance	67.8 %	67.8 %	67.0 %	75.3 %	47.1 %	50.6 %
	Tetranucleotide	99.8 %	99.8 %	99.9 %	99.8 %	100.0 %	100.0 %

– continued on next page

Table S1 – *continued from previous page*

Species		SPAdes	metaSPAdes	IDBA-UD	MetaVelvet	omega	mira
Pseudevernia furfuracea	Combined bins	2	2	2	4	4	4
	Removed contigs	33	69	58	117	85	125
	All Contigs	1,829	3,702	4,036	6,340	8,029	10,238
	Ascomycota contigs	1,216 (66 %)	1,808 (49 %)	2,195 (54 %)	3,457 (55 %)	4,995 (62 %)	4,174 (41 %)
	'Unknown' contigs	613	1,894	1,841	2,883	3,034	6,064
	Size	32.3 Mb	32.3 Mb	30.3 Mb	27.2 Mb	35 Mb	28.6 Mb
	N50	48,031	30,899	17,504	8,549	7,180	6,760
	GC	48.9 - 50.2 %	49.2 - 50.3 %	48.8 - 50.2 %	48.9 - 50.4 %	48.8 - 50.5 %	49.3 - 50.8 %
	tRNA's	22	15	20	15	18	13
	Coding density	55.0 %	56.6 %	56.0 %	58.0 %	56.6 %	58.3 %
	Coverage	152.4 x	153 x	153.7 x	152.4 x	129.4 x	110.8 x
	Profile completeness	54.8 %	54.5 %	48.3 %	51.5 %	52.9 %	50.0 %
	Profile copies	1.6 x	1.6 x	1.6 x	1.7 x	1.8 x	1.7 x
	Abundance	65.0 %	64.0 %	59.2 %	58.2 %	70.5 %	43.7 %
	Tetranucleotide	99.9 %	99.9 %	100.0 %	100.0 %	100.0 %	100.0 %

Table S2. Genome completeness for genome assemblies from metagenomic lichen thalli using different assemblers. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW). Percentage completeness is compared against 1,315 orthologous BUSCO marker genes for Ascomycota (C: complete [S: single-copy, D: duplicated], F: fragmented, M: missing).

(a) ***Evernia prunastri***

	unassigned assembly	metagenome MG	metagenome MW
SPAdes	C:95.5 % [S:94.1 %, D:1.4 %], F:2.7 %, M:1.8 %	C:95.4 % [S:95.0 %, D:0.4 %], F:3.2 %, M:1.4 %	C:93.2 % [S:92.9 %, D:0.3 %], F:3.1 %, M:3.7 %
metaSPAdes	C:95.5 % [S:94.1 %, D:1.4 %], F:2.7 %, M:1.8 %	C:95.2 % [S:94.8 %, D:0.4 %], F:3.2 %, M:1.6 %	C:93.7 % [S:93.4 %, D:0.3 %], F:3.1 %, M:3.2 %
IDBA-UD	C:94.4 % [S:93.2 %, D:1.2 %], F:3.0 %, M:2.6 %	C:93.8 % [S:93.5 %, D:0.3 %], F:3.4 %, M:2.8 %	C:91.3 % [S:91.0 %, D:0.3 %], F:3.2 %, M:5.5 %
metaVelvet	C:92.5 % [S:92.0 %, D:0.5 %], F:4.9 %, M:2.6 %	C:92.1 % [S:91.8 %, D:0.3 %], F:5.1 %, M:2.8 %	C:89.6 % [S:89.3 %, D:0.3 %], F:4.4 %, M:6.0 %
omega	C:51.1 % [S:50.7 %, D:0.4 %], F:33.5 %, M:15.4 %	C:49.2 % [S:49.2 %, D:0.0 %], F:34.2 %, M:16.6 %	C:36.7 % [S:36.7 %, D:0.0 %], F:24.8 %, M:38.5 %
mira	C:44.1 % [S:41.1 %, D:3.0 %], F:40.0 %, M:15.9 %	C:41.9 % [S:40.8 %, D:1.1 %], F:42.4 %, M:15.7 %	C:35.2 % [S:34.5 %, D:0.7 %], F:38.2 %, M:26.6 %

(b) ***Pseudevernia furfuracea***

	unassigned assembly	metagenome MG	metagenome MW
SPAdes	C:94.0 % [S:92.2 %, D:1.8 %], F:3.7 %, M:2.3 %	C:93.7 % [S:93.6 %, D:0.1 %], F:4.1 %, M:2.2 %	C:91.3 % [S:91.2 %, D:0.1 %], F:3.9 %, M:4.8 %
metaSPAdes	C:93.6 % [S:91.8 %, D:1.8 %], F:3.8 %, M:2.6 %	C:93.1 % [S:92.9 %, D:0.2 %], F:4.3 %, M:2.6 %	C:91.8 % [S:91.6 %, D:0.2 %], F:4.3 %, M:3.9 %
IDBA-UD	C:93.2 % [S:91.6 %, D:1.6 %], F:3.7 %, M:3.1 %	C:93.0 % [S:92.8 %, D:0.2 %], F:4.0 %, M:3.0 %	C:86.3 % [S:86.3 %, D:0.0 %], F:3.1 %, M:10.6 %
metaVelvet	C:83.3 % [S:81.6 %, D:1.7 %], F:10.6 %, M:6.1 %	C:82.2 % [S:81.3 %, D:0.9 %], F:10.9 %, M:6.9 %	C:78.6 % [S:77.9 %, D:0.7 %], F:7.9 %, M:13.5 %
omega	C:79.6 % [S:73.6 %, D:6.0 %], F:15.4 %, M:5.0 %	C:79.2 % [S:73.2 %, D:6.0 %], F:15.4 %, M:5.4 %	C:76.4 % [S:71.1 %, D:5.3 %], F:14.2 %, M:9.4 %
mira	C:70.2 % [S:66.2 %, D:4.0 %], F:22.6 %, M:7.2 %	C:69.8 % [S:66.3 %, D:3.5 %], F:22.7 %, M:7.5 %	C:58.3 % [S:56.4 %, D:1.9 %], F:17.9 %, M:23.8 %

Table S3. Number of contigs of assemblies based on metagenomic reads. Taxonomic assignment of metagenomic reads to Ascomycota was performed with MEGAN (MG) and MetaWatt (MW).

Species	metagenome assembly	SPAdes	metaSPAdes	IDBA-UD	MetaVelvet	omega	mira
Evernia prunastri	unfiltered and unassigned	51,428	67,447	98,046	16,779	34,042	148,894
	min 400 bp	49,746	61,771	95,085	14,558	32,679	130,534
	min 400 bp + Ascomycota MG	1,775	1,838	3,127	2,911	11,221	36,083
	min 400 bp + Ascomycota MW	1,624	1,602	2,757	2,432	7,577	39,338
	overlap (identical contigs in MG and MW)	1,026	1,275	1,527	2,076	5,647	21,409
Pseudevernia furfuracea	unfiltered and unassigned	62,248	101,711	145,738	226,304	26,661	143,739
	min 400 bp	60,165	90,678	144,848	126,609	26,130	113,825
	min 400 bp + Ascomycota MG	3,558	4,794	6,006	8,384	10,229	21,159
	min 400 bp + Ascomycota MW	1,829	3,702	4,036	6,340	8,029	10,238
	overlap (identical contigs in MG and MW)	1,300	2,190	2,516	3,997	5,838	6,831

Table S4. Taxonomic assignment of quality filtered metagenomic reads from lichen thalli that were not assigned as reference lichen-forming fungus (see Fig. 4 in the main text). These reads were classified using DIAMOND/MEGAN against the NCBI GenBank *nr* protein database. Only phyla above 0.1 % are presented.

		Evernia prunastri		***Pseudvernia furfuracea***	
		7,880,997	26.6 %	9,776,595	27.8 %
No blast hits		4,176,471	14.1 %	7,863,588	22.4 %
Not assigned		16,915	0.1 %	5,618	0.0 %
Archaea		442	0.0 %	482	0.0 %
Bacteria		3,168,381	10.7 %	659,705	1.9 %
Eukaryota		457,096	1.5 %	1,197,044	3.4 %
Kingdom	Fungi	213,931	0.7 %	346,484	1.0 %
	Metazoa	17,126	0.1 %	27,789	0.1 %
	Viridiplantae	189,160	0.6 %	743,704	2.1 %
Phylum (> 0.1 %)	Acidobacteria	415,805	1.4 %	197,181	0.6 %
	Actinobacteria	74,317	0.3 %	30,936	0.1 %
	Ascomycota	154,470	0.5 %	302,340	0.9 %
	Basidiomycota	49,305	0.2 %	31,477	0.1 %
	Chlorophyta	131,863	0.4 %	573,746	1.6 %
	Proteobacteria	2,445,225	8.3 %	307,007	0.9 %
	Streptophyta	38,053	0.1 %	97,982	0.3 %

Table S5. Taxonomic assignment of scaffolds assembled with SPAdes from metagenomic reads from lichen thalli. These scaffolds were classified using DIAMOND/MEGAN with the NCBI GenBank *nr* protein database. Only phyla above 0.1 % are presented.

			Evernia prunastri				*Pseudevernia furfuracea*			
			number of scaffolds		length		number of scaffolds		length	
Total			49,746		126,470,248		60,165		186,849,783	
No blast hits			15,751	31.7 %	19,909,906	15.7 %	23,231	38.6 %	31,266,115	16.7 %
Not assigned			6,157	12.4 %	9,350,115	7.4 %	9,401	15.6 %	20,532,506	11.0 %
Eukaryota			8,420	16.9 %	50,879,619	40.2 %	20,748	34.5 %	117,884,326	63.1 %
	Fungi		2,296	4.6 %	39,589,067	31.3 %	3,880	6.4 %	43,657,327	23.4 %
		Ascomycota	1,775	3.6 %	38,671,927	30.6 %	3,558	5.9 %	42,531,237	22.8 %
		Basidiomycota	407	0.8 %	490,161	0.4 %	193	0.3 %	394,925	0.2 %
	Viridiplantae		5,101	10.3 %	9,608,042	7.6 %	15,683	26.1 %	67,410,521	36.1 %
		Chlorophyta	3,333	6.7 %	6,443,731	5.1 %	13,013	21.6 %	49,318,344	26.4 %
		Streptophyta	902	1.8 %	1,623,011	1.3 %	1,275	2.1 %	8,306,445	4.4 %
	Metazoa		201	0.4 %	365,771	0.3 %	241	0.4 %	1,439,461	0.8 %
		Arthropoda	25	0.1 %	51,029	0.0 %	32	0.1 %	164,902	0.1 %
		Chordata	75	0.2 %	147,033	0.1 %	83	0.1 %	547,521	0.3 %
Bacteria			19,058	38.3 %	45,660,763	36.1 %	6,335	10.5 %	14,666,264	7.8 %
		Acidobacteria	3,042	6.1 %	10,455,677	8.3 %	3,415	5.7 %	6,662,926	3.6 %
		Actinobacteria	479	1.0 %	494,543	0.4 %	57	0.1 %	114,791	0.1 %
		Bacteroidetes	35	0.1 %	56,628	0.0 %	21	0.0 %	139,474	0.1 %
		Cyanobacteria	57	0.1 %	105,054	0.1 %	47	0.1 %	222,491	0.1 %
		Firmicutes	50	0.1 %	93,783	0.1 %	45	0.1 %	150,082	0.1 %
		Proteobacteria	14,109	28.4 %	31,900,387	25.2 %	2,072	3.4 %	5,241,637	2.8 %
Viruses			16	0.0 %	22,036	0.0 %	53	0.1 %	299,184	0.2 %

A.2. Biosynthetic gene clusters in two closely related sister-species

Statement of author contributions

Title: The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo

Journal: The Lichenologist 50(3), 329–40, doi:10.1017/S002428291800021X

Status: Accepted on 7 January 2018

Authors: Francesco Dal Grande (FDG), Anjuli Meiser (AM), Bastian Greshake Tzovaras (BGT), Jürgen Otte (JO), Ingo Ebersberger (IE), Imke Schmitt (IS)

Author contributions:

(1) Development and planning	
Doctoral candidate (AM):	30 %
Coauthor FDG:	60 %
Coauthor IS:	10 %

(2) Conduction of experiments	
Doctoral candidate (AM):	25 % Sample collection for *U. pustulata* RNA, taxonomic assignments
Coauthor FDG:	50 % Sample collection
Coauthor JO:	25 % DNA isolation, sequencing library preparation

(3) Compilation of data sets and figures/tables	
Doctoral candidate (AM):	30 % Secondary metabolite gene cluster annotation, preparation of related table, genome annotation
Coauthor FDG:	60 % Data quality trimming, genome assemblies, all figures/tables
Coauthor BGT:	10 % Taxonomic assignment, genome annotation

(4) Data analyses and interpretation	
Doctoral candidate (AM):	30 % Taxonomic assignment, gene annotation, orthologous gene sets, genome mining
Coauthor FDG:	60 % Synteny analysis, comparison of gene sets, data interpretation
Coauthor IS:	10 % Data interpretation

(5) Preparation of manuscript	
Doctoral candidate (AM):	20 %
Coauthor FDG:	60 %
Coauthor BGT:	10 %
Coauthor IE:	5 %
Coauthor IS:	5 %

The Lichenologist **50**(3): 329–340 (2018) doi:10.1017/S002428291800021X

The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo

Francesco DAL GRANDE, Anjuli MEISER, Bastian GRESHAKE TZOVARAS, Jürgen OTTE, Ingo EBERSBERGER and Imke SCHMITT

Abstract: *Lasallia hispanic*a (Frey) Sancho & A. Crespo is one of three *Lasallia* species occurring in central-western Europe. It is an orophytic, photophilous Mediterranean endemic which is sympatric with the closely related, widely distributed, highly clonal sister taxon *L. pustulata* in the supra- and oro-Mediterranean belts. We sequenced the genome of *L. hispanica* from a multispore isolate. The total genome length is 41·2 Mb, including 8488 gene models. We present the annotation of a variety of genes that are involved in protein secretion, mating processes and secondary metabolism, and we report transposable elements. Additionally, we compared the genome of *L. hispanica* to the closely related, yet ecologically distant, *L. pustulata* and found high synteny in gene content and order. The newly assembled and annotated *L. hispanica* genome represents a useful resource for future investigations into niche differentiation, speciation and microevolution in *L. hispanica* and other members of the genus.

Key words: functional annotation, mating type, polyketide synthase, secretome, synteny, transposable elements

Accepted for publication 7 *January* 2017

Introduction

Lasallia hispanica (Frey) Sancho & A. Crespo represents one of three species of the genus *Lasallia* which occur in central Western Europe (Sancho & Crespo 1989). The three *Lasallia* species differ in distribution, habitat preference, morphology and mode of reproduction. *Lasallia pustulata* has the widest distribution, occurring in Mediterranean to boreal-montane habitats from southern Europe to northern Scandinavia (Hestmark 1992; Rolshausen *et al.* 2018). The other two congeners are endemic to the Mediterranean region: *L. hispanica* prefers supra- and oro-Mediterranean habitats in the Iberian Peninsula, southern Italy and northern Morocco, and *L. brigantium* is confined to coastal areas in west Corsica and north-west Sardinia below 300 m a.s.l. (Sancho & Crespo 1989). *Lasallia hispanica* is sympatric with *L. pustulata* in the supra- and oro-Mediterranean bioclimatic belts (Sancho & Crespo 1989) where the two species often share the same photobiont (Dal Grande *et al.* 2017). *Lasallia hispanica* and *L. pustulata* differ in their water acquisition strategies: *L. pustulata* relies on surface run-offs, whereas *L. hispanica* takes up moisture directly from fog and low-lying clouds, therefore becoming desiccated more rapidly and more frequently (Vivas *et al.* 2017). A recent study comparing the photosynthetic performance of the two species in nature and under laboratory conditions suggests that *L. hispanica* might be more

F. Dal Grande, A. Meiser, J. Otte, I. Ebersberger and I. Schmitt: Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.
B. Greshake Tzovaras and I. Ebersberger: Institute of Cell Biology and Neuroscience, Goethe University Frankfurt, Max-von-Laue Str. 13, 60438 Frankfurt am Main, Germany.
A. Meiser and I. Schmitt: Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.
B. Greshake Tzovaras: Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
Current address for F. Dal Grande (corresponding author): Departamento de Farmacología, Farmacognosia y Botánica, Facultad de Farmacia, Universidad Complutense de Madrid, 28040 Madrid, Spain. Email: fdalgran@ucm.es

resistant to environmental stress than *L. pustulata*. This is probably due to the more efficient and rapid activation of stress-related repair mechanisms in *L. hispanica* (Vivas *et al*. 2017). The three *Lasallia* species have a mixed asexual and sexual reproductive strategy. However, reproduction in *L. pustulata* is predominantly vegetative, by means of isidia, while *L. hispanica* and *L. brigantium* predominantly reproduce sexually (Sancho & Crespo 1989). While *L. pustulata* has been used as a model to explore climate adaptation in lichens (Dal Grande *et al*. 2018) and symbiont-driven ecological expansion (Rolshausen *et al*. 2018), molecular studies on *L. hispanica* are lacking. The genetic differentiation among the three species has yet to be explored.

The genomics revolution is transforming the way we study evolution and ecology (Wolfe & Li 2003; Grube *et al*. 2014). Evolutionary genomics and phylogenomics further our understanding of speciation, phylogenetic relationships and the evolutionary origin of functional traits in lichenized fungi. Phylogenomic datasets have been used to resolve evolutionary relationships in the *Rhizoplaca melanophthalma* species complex (Chan & Ragan 2013; Leavitt *et al*. 2016). Comparative genomics has been used to reveal gene family size changes and gene deletions associated with lichenization in *Endocarpon pusillum* (Wang *et al*. 2014), to derive phylogenetic markers useful for resolving relationships among close relatives (Magain *et al*. 2017), and to study the properties and evolution of mitochondrial genomes (Xavier *et al*. 2012).

Ecological genomics is an emerging field in lichenology. It allows questions to be addressed related to, for example, niche differentiation, ecological specialization and local adaptation. Transcriptomics has been employed to infer the response of *Peltigera membranacea* and its cyanobiont to thermal stress (Steinhäuser *et al*. 2016), and of *Trebouxia* to desiccation (Candotto Carniel *et al*. 2016). Recently, we used a population genomics approach based on whole-genome resequencing of pools of DNA from lichen populations to study the genomic signatures of adaptation in *L. pustulata* along an altitudinal gradient (Dal Grande *et al*. 2017). In this study we revealed the existence of two locally adapted ecotypes using correlations between single-nucleotide polymorphisms (SNPs) and environmental parameters.

Lichen metagenomics (i.e. the direct sequencing of mixed genomic material from lichen thalli) represents a cultivation-independent approach to explore the diversity and functional aspects of the lichen symbiosis. For instance, it is possible to reconstruct the genomes of the individual symbiotic partners using a single, short-read sequencing library layout (i.e. metagenome skimming; Greshake Tsovaras *et al*. 2016; Meiser *et al*. 2017). Metagenomic lichen samples have also been used to apply restriction site-associated DNA sequencing (RAD-seq) for phylogenetic reconstructions of lichenized fungi based on genomic sequence information (Grewe *et al*. 2017). Genome mining is increasingly employed to survey lichens for genes associated with the biosynthesis of active metabolites, revealing in some cases unexpected biosynthetic potential (e.g. Kampa *et al*. 2013). For example, *Cladonia uncialis* contained a gene cluster responsible for the biosynthesis of a halogenated isocoumarin (Abdel-Hameed *et al*. 2016). The advent of long-read sequencing technologies from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies will drastically improve the assembly process as well as the *in-silico* separation of organisms from mixed DNA samples.

Here we present the *de novo* assembly and annotation of the genome of *L. hispanica*. Using Illumina next-generation sequencing technology we obtained and annotated a high-quality draft genome. We identified gene clusters associated with secondary metabolite biosynthesis, mating-type loci and transposable elements, and compared them to the closely related *L. pustulata* (Davydov *et al*. 2010). Finally, we established synteny and orthology between *L. hispanica* and *L. pustulata*. In addition to providing structured data for various phylogenetic studies, the work presented here will provide a genomic resource for further studies aiming to 1) understand the basis of polygenic adaptation in *L. hispanica* based on

population genomic resequencing of natural populations, 2) study the impact of different reproductive strategies on the evolution of genomes and populations in *L. hispanica* and *L. pustulata*, and 3) infer the genomic footprints of niche differentiation of the two species.

Materials and Methods

In vitro cultivation of the lichen-forming fungus *Lasallia hispanica*

The lichen-forming fungus *L. hispanica* was isolated *in vitro* from a specimen collected from Puerto de Pico (Ávila, Spain; 40·322527°, −5·013808°, 1350 m a.s.l.; hb. Senckenbergianum voucher no. FR-0265086) in June 2014. The mycobiont culture (Schmitt laboratory, SBiK-F, C0002) was obtained from a multispore discharge from a single apothecium of *L. hispanica* following the method of Yamamoto *et al.* (1985). Briefly, apothecia were picked from the thallus, washed under distilled running water for several minutes and transferred individually onto inverted 4% water agar plates with sterile nylon membrane filters for 48 h. After ejection, the filters with the spores were transferred to germination medium in Petri dishes (Denison 2003). Upon germination, the spores were transferred to malt yeast extract medium. The mycobiont colonies were maintained at room temperature in darkness and were sub-cultured monthly onto fresh medium until sufficient biomass for genomic analysis was obtained (*c.* 6 months; Fig. 1).

DNA isolation and sequencing

About 0·5 g of mycobiont mycelia was collected and ground in liquid nitrogen with a mortar and pestle. Genomic DNA was isolated using the CTAB Maxi-prep method (Cubero & Crespo 2002), resulting in a total yield of *c.* 5 µg DNA. Three Illumina genomic libraries were sequenced: 1) short-insert DNA library, paired-end (300 bp), on Illumina MiSeq, 2) Nextera mate-pair library with 3 kb inserts, 3) Nextera mate-pair library with 8 kb inserts. Sequencing was performed at StarSeq (Mainz, Germany).

Genome assembly and annotation

Adapters and low quality short-insert reads were trimmed (i.e. Q score < 20 in a sliding window of 5 bp, minimum length < 100 bp) using Trimmomatic 0.36 (Bolger *et al.* 2014). The reads were further quality-filtered using the software Sickle v.1.33 (*-l 127 -q 20*;

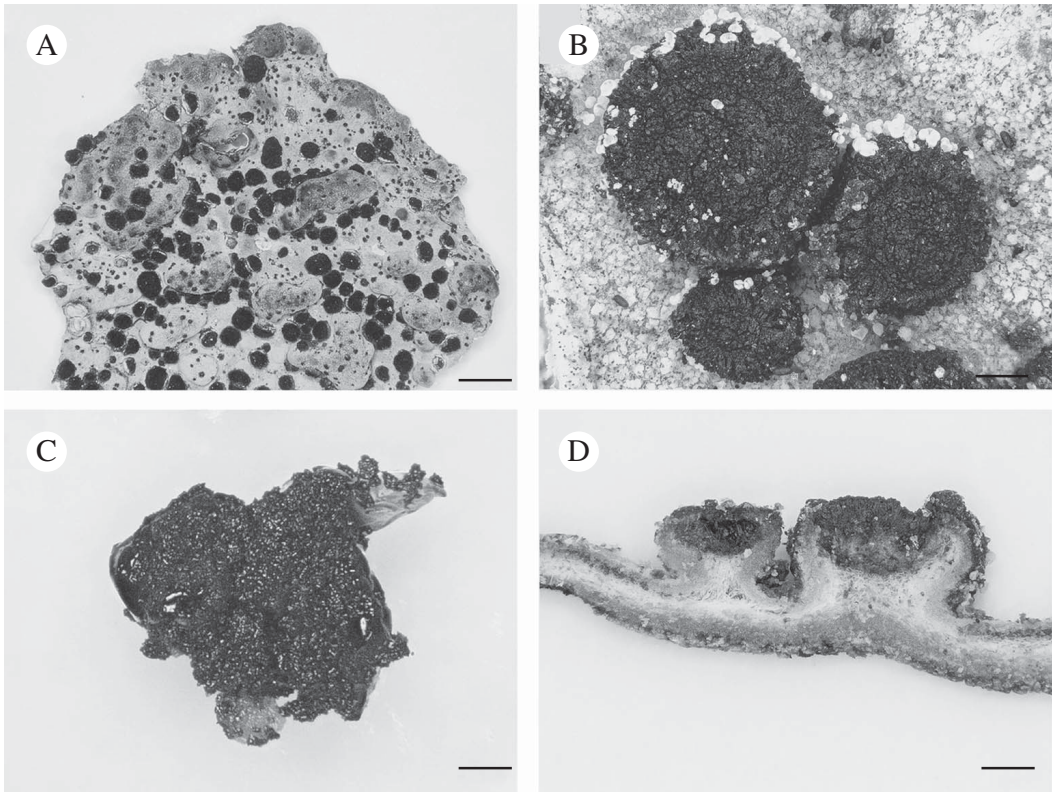


FIG. 1. *Lasallia hispanica*. A, thallus with apothecia; B, apothecia; C, mycobiont culture used for genome sequencing; D, section of thallus with apothecia. Scales: A = 10 mm; B–D = 1 mm.

available at https://github.com/najoshi/sickle). Adapters were removed from the mate-pair reads using NxTrim v.0.3.2 (O'Connell *et al.* 2015). Prior to genome assembly, we assembled overlapping pairs of short-insert reads using PEAR v.0.9.6 (Zhang *et al.* 2014). Reads were subsequently assembled *de novo* using SPAdes v.3.9.0 (*-k 21,33,55,77,99,127*; Bankevich *et al.* 2012).

We filled gaps between contigs using SSPACE (Boetzer *et al.* 2011) and GapFiller (Boetzer & Pirovano 2012). To filter the assembly from potential contaminants and to extract contigs of fungal origin, we taxonomically assigned the scaffolds using MetaWatt v.3.5.3 (Strous *et al.* 2012) against a non-redundant database consisting of genomes from 122 Archaea, 1747 Bacteria, 514 Eukaryota and 535 Viruses. We estimated genome completeness of the newly assembled *L. hispanica* genome using BUSCO v.2.0 (Benchmarking Universal Single-Copy Orthologs; Simão *et al.* 2015) and a lineage-specific set of Ascomycota single-copy orthologs.

The newly assembled genome of *L. hispanica* was annotated using funannotate v.0.5.4 (https://github.com/nextgenusfs/funannotate). As training data for funannotate, RNAseq data from *L. pustulata* (Dal Grande *et al.* 2017) was assembled using Trinity and PASA and used along the unassembled reads. Furthermore, we used the predicted protein sequences from *Xanthoria parietina* (https://genome.jgi.doe.gov/Xanpa2/Xanpa2.home.html) and *Cladonia grayi* (https://genome.jgi.doe.gov/Clagr3/Clagr3.home.html) as training data for the gene prediction. Blast2GO v.4.1.9 (Conesa *et al.* 2005) was used to annotate the predicted protein sequences with gene ontology (GO) terms and protein names using the NCBI nr database at an E-value cut-off of 1×10^{-3} and default weighting parameters. The functional annotations were simplified to a set of broad terms by mapping the GO annotations to the Generic GO-Slim terms using Blast2GO.

Repeat elements

We surveyed the draft genome of *L. hispanica* for transposable elements (TEs) and repeated sequences. For this purpose, we first constructed a reference TE consensus library using the TEdenovo (Flutre *et al.* 2011; Hoede *et al.* 2014) and the TEannot (Quesneville *et al.* 2005) from the REPET TE annotation pipelines for the high quality PacBio assembly of the *L. pustulata* genome. These sequences were used as probes to annotate the *L. hispanica* genome with TEannot from the REPET pipeline. TE consensus nucleotide sequences were classified according to the Repbase database (Jurka *et al.* 2005) and named according to the classification proposed by Wicker *et al.* (2007).

Secreted proteins

To identify proteins with an extracellular secretion signal, we used SignalP v.4.0 (Petersen *et al.* 2011), TargetP v.1 (Emanuelsson *et al.* 2007) and Tmhmm2.0c (Krogh *et al.* 2001). Only annotated protein-coding genes having a signal peptide and not having a membrane localization domain were considered as putatively secreted.

Mating-type annotation

MAT alleles are typically flanked by the putative DNA lyase (*APN2*) and the cytoskeleton assembly control (*SLA2*) genes (Debuchy & Turgeon 2006). We identified the *MAT* locus in *L. hispanica* and *L. pustulata* using BlastP searches against a database composed of *ADN2*, *SLA2*, *MAT1-1*, and *MAT1-2* protein sequences of various ascomycetes, including lichen-forming fungi.

Annotation of genes and gene clusters associated with secondary metabolite biosynthesis

Genes and gene clusters involved in secondary metabolism in *L. hispanica* and *L. pustulata* were predicted using antiSMASH fungal v.4.0.0 (fungiSMASH; Blin *et al.* 2017).

Synteny and orthology analysis

We compared the genome of the closely related species *L. pustulata* (Greshake Tsovaras 2018) to find orthologous gene pairs between the two species. For this purpose, we identified reciprocal best BLAST hits (RBH) between the two gene sets. This approach constitutes a relatively simple and fast method for finding orthologs between different assemblies of the same or closely related species (Ward & Moreno-Hagelsieb 2014). We ran BLAST v.2.2.30+ using Smith-Waterman alignment and soft filtering (*use_sw_tback*, *soft_masking true*, *seq yes*, *evalue 1e-6*) for better detecting orthologs as RBH (Moreno-Hagelsieb & Latimer 2008; Ward & Moreno-Hagelsieb 2014). To identify RBH we filtered the BLAST output for a minimum identity of 70% over the alignment length and a minimum query coverage of 50% (Camacho *et al.* 2009), sorted for the highest bitscore and lowest E-value, and manually removed multiple identical top hits, if present.

Lasallia hispanica and *L. pustulata* assemblies and gene sets were compared to identify genomic portions in which gene order is conserved (i.e. syntenic regions). For this purpose, we used SyMap v.4.2 (Synteny Mapping and Analysis Program; Soderlund *et al.* 2011) to compute and display syntenic relationships between *L. hispanica* and *L. pustulata*. For this, we aligned scaffolds longer than 50 kb of each species using MUMmer (Kurtz *et al.* 2004) and used synteny to order the draft genome (*L. hispanica*) against the reference (*L. pustulata*). To calculate the percentage of genes located in syntenic blocks, gene coordinates of the two species were imported into SyMap as *.gff*.

Results and Discussion

Genome assembly and annotation

After adapter removal, and length and quality filtering, we obtained 11 313 695 short-insert paired-end reads, plus 3 163 139 and 3 351 197 mate pair reads for the 3 kb and 8 kb libraries, respectively. These reads

TABLE 1. *Information on the* L. hispanica *genome assembly.*

Scaffolds		Genes	
Total number	1619	Total number	8488
Total size (bp)	41 207 996	Proportion covered by genes (%)	33
Longest scaffold (bp)	615 827	Mean protein size in aa (min/max)	470 (50/6, 195)
Mean size (bp)	25 453	GC content (%)	51·2
Median size (bp)	3294	Coding region GC (%)	54·1
N50 length (bp)	145 035		

were assembled using SPAdes into 1619 scaffolds longer than 500 bp (N50 = 145 035; Table 1). The draft assembly has a total length of 41·2 Mb and a coverage of approximately 160×. The evaluation of the genome completeness of our draft genome assembly based on 1315 single-copy fungal orthologs showed that most of the gene space was covered (96·3%). The *L. hispanica* genome assembly contained 1256 complete and single-copy, 10 duplicated, 27 fragmented and 22 missing BUSCO genes. The overall GC content of the *L. hispanica* genome is 51·2%. The GC content of gene coding sequences increases to 54·1% and is similar to that of *L. pustulata* (overall GC = 51·7%; CDS GC = 53·2%).

We predicted a total of 8488 *ab initio* gene models, of which 3929 (46·3%) were assigned a total of 15 820 GO terms. The most abundant biological process GO-Slim terms were organic substance metabolic process (15·6%), cellular metabolic process (15·2%), primary metabolic process (14·7%) and nitrogen compound metabolic process (10·6%). Abundant molecular function GO-Slim terms included organic cyclic compound binding (17·8%), ion binding (15·6%), hydrolase activity (11·7%) and transferase activity (11·2%). Finally, most of the cellular components GO-Slim terms were categorized as intracellular (19·9%), intracellular part (19·4%), intracellular organelle (15·3%) and membrane-bounded organelle (13·2%) (Fig. 2).

Transposable elements

Transposable Elements (TEs) are DNA fragments with the ability to move within the genome by generating new copies of themselves. TEs are an important source of mutations in genomes and may promote genome restructuring and chromosome instability due to their repeated nature (Bonchev & Parisod 2013). TEs are typically divided into two classes depending on their mechanism of mobility: retrotransposons (class I) and DNA transposons (class II) (Wicker *et al.* 2007). The cut-and-paste transposition mechanism of retrotransposons involves an RNA intermediate which is reverse transcribed by a reverse transcriptase often encoded by the TE itself. DNA transposons instead transpose directly from DNA to DNA.

In fungi, 0–30% of the genome consists of transposable elements, with LTR (Long Terminal Repeats)-retrotransposons usually representing the largest fraction (Castanera *et al.* 2016). The repetitive nature of TE sequences, in combination with short-read sequencing technologies, exacerbates the correct assembly of TEs, especially for TE families exhibiting high sequence identity, high copy number or complex genomic arrangements (Nilsson 2016).

Transposable elements were found to cover 21·23% of the *L. pustulata* genome for a total of *c.* 7 Mbp, including 70 class I and 35 class II elements with full length copies (444–11 000 bp, mean size: 4021 bp) (see Supplementary Material Table S1, available online). Conversely, the draft genome of *L. hispanica* displayed an almost complete absence of full length elements. These results confirm the limitation of the short-read sequencing technology in reconstructing TEs. Therefore, the current resolution of this draft genome, like most Illumina-based genome assemblies, is insufficient to give a detailed picture of the TE content.

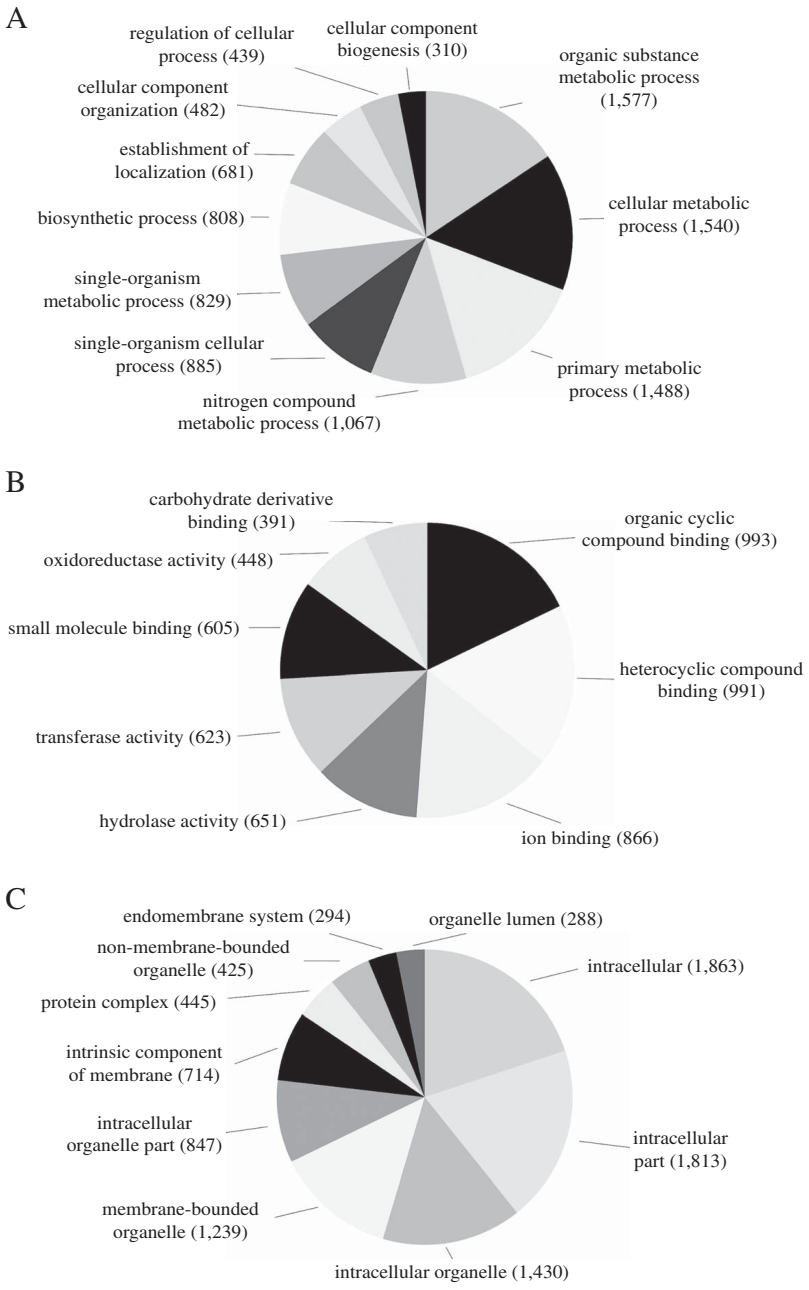


FIG. 2. Distribution of Blast2GO annotations for *L. hispanica*. Charts show level 3 annotations for Biological Process (A), Molecular Function (B) and Cellular Components (C).

Secreted proteins

The secretion of proteins and other enzymes into the extracellular environment is a vital process in fungi (Krijger *et al.* 2014). In particular, secreted proteins play an essential role in nutrient acquisition and self-protection. Furthermore, the fungal secretome directly or indirectly modulates interactions of the fungus with living and non-living substrata, including recognition processes (Wessels 1993). We found 104

genes encoding putatively secreted proteins in *L. hispanica*, including 16 glycoside hydrolases, six carboxipeptidases and two glucoamylases. Putatively secreted proteins ranged in length from 61 to 1672 aa (see Supplementary Material Table S2, available online).

Mating types

The mating system of filamentous ascomycetes is usually represented by one locus (i.e. the *MAT* locus) which encodes proteins of the high-mobility-group (HMG) superfamily (Coppin *et al.* 1997). The *MAT* locus is typically present in two complementary forms (i.e. idiomorphs) referred to as *MAT1-1* and *MAT1-2* (or *MAT-1* and *MAT-2*). Homothallic species typically contain both *MAT* genes (i.e. *MAT-1* encoding a protein with a MATα_HMG domain and *MAT-2* encoding a protein with a MATA_HMG domain) within the same genome. Heterothallic species instead contain a single *MAT* locus; isolates can thus carry either *MAT-1* or *MAT-2* genes (Kronstad & Staben 1997). In this study we identified the *MAT* loci in the *L. hispanica* and *L. pustulata* genomes.

Only one complete mating-type locus was found in the genome assembly for *L. hispanica*: *MAT1-2* containing the MATA_HMG domain. The orthologous *MAT1-2* idiomorph was also found in a newly assembled genome of *L. pustulata* (Greshake Tsovaras 2018). As in *L. hispanica*, the *MAT1-2* idiomorph of *L. pustulata* includes an unknown gene containing a homeodomain. The complementary mating idiomorph (i.e. *MAT1-1*) was also found in our first draft assembly of *L. pustulata* available at the European Nucleotide Archive GCA_000938525.1 obtained from a different thallus. This region lacks *MAT1-2* and the homeodomain-containing gene, while it includes a full *MAT1-1* gene with the MATα_HMG (Fig. 3). Our results provide evidence for a heterothallic lifestyle of both *Lasallia* species. However, inferences based on genome sequence analysis require additional experimental validation, including analysis of single-spore isolates and estimation of *MAT* frequencies in natural populations using *MAT*-idiomorph specific probes (Honegger *et al.* 2004; Singh *et al.* 2012, 2015; Alors *et al.* 2017; Ludwig *et al.* 2017).

Secondary metabolite biosynthetic genes and gene clusters

The advent of genome sequencing technologies is revolutionizing the field of natural product discovery (Doroghazi *et al.* 2014). Whole-genome mining of biosynthetic gene clusters has revealed a large number of uncharacterized secondary metabolite gene clusters in various organisms, including lichen-forming fungi (e.g. Kampa *et al.* 2013; Abdel-Hameed *et al.* 2016).

HPLC analyses revealed similarities in the chemical profiles of *L. hispanica* and *L. pustulata*, with gyrophoric acid as the major compound and traces of lecanoric, umbilicaric, hiascic acids and skyrin (Posner *et al.* 1991). In the *L. hispanica* genome we identified 18 secondary metabolite clusters with complete core biosynthetic genes (core biosynthetic genes = polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS),

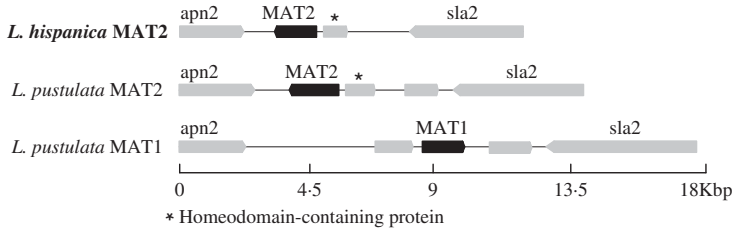


FIG. 3. Configuration of the *MAT* loci in *L. hispanica* and in two *L. pustulata* assemblies (*MAT1*: European Nucleotide Archive GCA_000938525.1; *MAT2*: Greshake (2018)).

TABLE 2. *Biosynthetic genes and gene clusters in* L. hispanica *and* L. pustulata. *A dash indicates no genes were detected for that class.*

Class	*L. hispanica*	*L. pustulata*	Core biosynthetic gene orthologs	Ortholog gene clusters
Non-reducing PKS	5	7	5	4
Reducing PKS	6	6	3	2
Hybrid non-red/red PKS	2	–	–	–
Hybrid PKS_NRPS	–	1	–	–
Type III PKS	1	1	1	1
Terpene synthase	4	5	4	4
Lantipeptide synthetase	–	1	–	–
Partial PKS	3	3	–	–

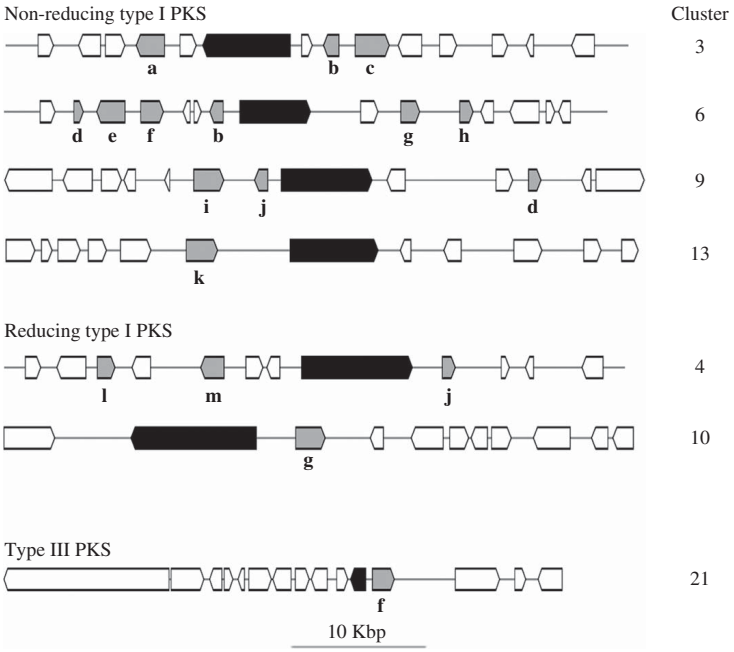


FIG. 4. Configuration of *L. hispanica* biosynthetic gene clusters with orthologs in *L. pustulata*. Black boxes represent core biosynthetic genes (PKSs in the upper six clusters and a chalcone and stilbene synthase in the bottom cluster). Shaded boxes indicate genes coding for tailoring enzymes: a, acyltransferase; b, metallo-beta-lactamase family protein; c, halogenase; d, aldo/keto reductase; e, drug resistance transporter EmrB/QacA; f, cytochrome P450; g, *O*-methyltransferase; h, haloalkane dehalogenase; i, dioxygenase TauD/TfdA; j, FAD-linked oxidase domain protein; k, serine/threonine protein kinase; l, acyl-CoA dehydrogenase; m, AMP-dependent synthetase and ligase.

etc.) (Table 2, Supplementary Material Table S3). Among the non-reducing type I PKS, three genes showed duplicated ACP domains (Supplementary Material Table S4, available online). Interestingly, we found only partial homology between the biosynthetic gene clusters of *L. hispanica* and *L. pustulata*, with 13 putative orthologs among 40 complete, core biosynthetic genes of the two species (Table 2, Supplementary Material Table S3). Eleven biosynthetic clusters, including four non-reducing and two reducing PKS, four terpene synthases and one type III PKS, showed high similarity of core genes and genes coding for tailoring enzymes. These clusters therefore represent ideal candidates for the biosynthesis of natural compounds that are shared between the two lichen species (Fig. 4).

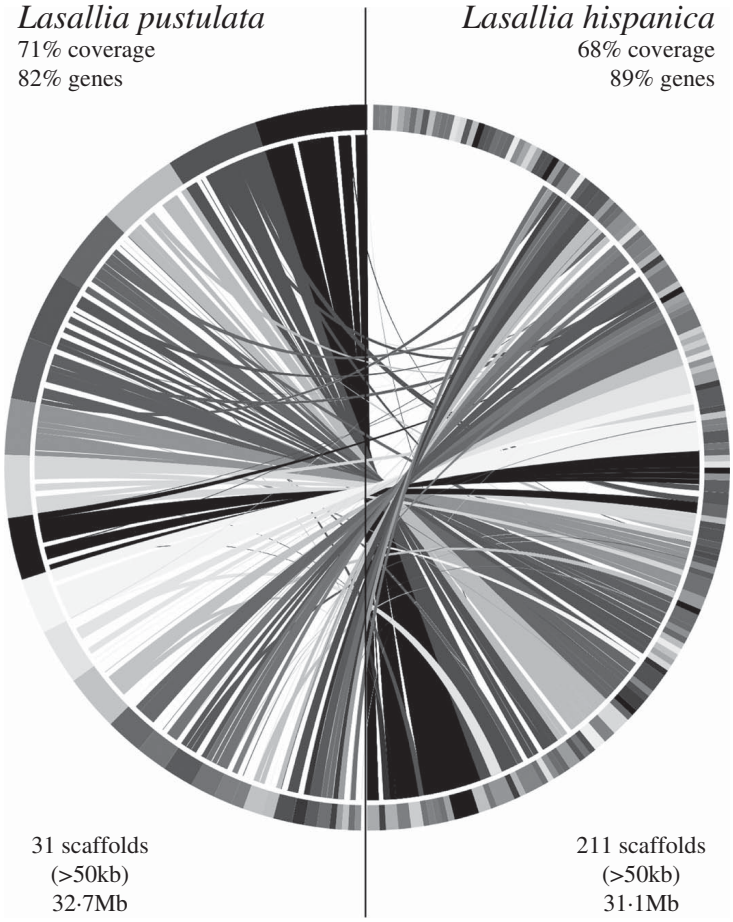


FIG. 5. Circle plot of the genome alignment between 31 *L. pustulata* (left) and 211 *L. hispanica* (right) scaffolds. Scaffolds of *L. hispanica* were ordered to align against the genome of *L. pustulata* using information from 202 syntenic blocks.

Our results suggest that both *Lasallia* species have a far greater potential to produce specialized secondary metabolites than previously thought. Genomics-driven discovery of fungal natural products and comparison of gene clusters between closely related species with similar chemical profiles is just the first step towards linking these gene clusters to their metabolites (Chooi & Solomon 2014).

Synteny and orthology analysis

Based on RBH analysis, 6382 orthologous gene pairs were identified between *L. hispanica* and *L. pustulata* proteins (see Supplementary Material Table S5, available online). The 211 largest (i.e. >50 kb) *L. hispanica* scaffolds (representing 75·6% of the genome) were then aligned with the 31 largest *L. pustulata* scaffolds (99·5% of the genome) to find syntenic regions. The alignment produced 68% and 71% of syntenic coverage in *L. hispanica* and *L. pustulata*, respectively, with gene retention >80% for both species. The circle plot of this genome comparison shows a high degree of synteny conservation between *L. hispanica* and *L. pustulata*, with only a few rearrangements (Fig. 5).

The draft genome of *L. hispanica* presented in this study sets the foundation for further research into speciation and niche evolution mechanisms in lichen-forming fungi. We believe that the *L. hispanica*-*L. pustulata* system is

particularly suitable for this application owing to the ecological, reproductive and genetic differences between the species. In addition, the annotated draft genome serves as a resource for developing molecular markers, targeting specific functional genes and analysing repetitive elements in the context of population studies.

Data accessibility

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the Accession PKMA00000000. The raw sequence reads are available under the Accession number SRP127347.

This manuscript is dedicated to our friend and mentor Ana Crespo on the occasion of her 70th birthday. We honour her invaluable contributions to lichenology.

We thank Pradeep K. Divakar and Ana Crespo (Madrid) for support with fieldwork, and Véronique Jamilloux, Nathalie Choisne and Joelle Amselem (INRA - URGI, Versailles) for training and support in the use of REPET. Some analyses were performed on the FUCHS cluster of the Center for Scientific Computing (CSC) at Goethe University in Frankfurt am Main. This research was funded by Landes-Offensive zur Entwicklung Wissenschaftlich-Oekonomischer Exzellenz (LOEWE) of Hesse's Ministry of Higher Education, Research and the Arts through the Senckenberg Biodiversity and Climate Research Centre (SBiK-F).

Supplementary Material

For supplementary material accompanying this paper visit https://doi.org/10.1017/S002428291800021X

References

Abdel-Hameed, M., Bertrand, R. L., Piercey-Normore, M. D. & Sorensen, J. L. (2016) Identification of 6-hydroxymellein synthase and accessory genes in the lichen *Cladonia uncialis*. *Journal of Natural Products* **79:** 1645–1650.

Alors, D., Dal Grande, F., Cubas, P., Crespo, A., Schmitt, I., Molina, M. C. & Divakar, P. K. (2017) Panmixia and dispersal from the Mediterranean Basin to Macaronesian Islands of a macrolichen species. *Scientific Reports* **7:** 40879.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19:** 455–477.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., *et al.* (2017) antiSMASH 4.0–improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* **1:** 1–6.

Boetzer, M. & Pirovano, W. (2012) Toward almost closed genomes with GapFiller. *Genome Biology* **13:** R56.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27:** 578–579.

Bolger, A. M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120.

Bonchev, G. & Parisod, C. (2013) Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources* **13:** 765–775.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Candotto Carniel, F., Gerdol, M., Montagner, A., Banchi, E., De Moro, G., Manfrin, C., Muggia, L., Pallavicini, A. & Tretiach, M. (2016) New features of desiccation tolerance in the lichen photobiont *Trebouxia gelatinosa* are revealed by a transcriptomic approach. *Plant Molecular Biology* **91:** 319–339.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., *et al.* (2016) Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics* **12:** 1–27.

Chan, C. X. & Ragan, M. A. (2013) Next-generation phylogenomics. *Biology Direct* **8:** 3.

Chooi, Y.-H. & Solomon, P. S. (2014) A chemical ecogenomics approach to understand the roles of secondary metabolites in fungal cereal pathogens. *Frontiers in Microbiology* **5:** 1–7.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. & Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21:** 3674–3676.

Coppin, E., Debuchy, R., Arnaise, S. & Picard, M. (1997) Mating types and sexual development in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* **61:** 411–428.

Cubero, O. F. & Crespo, A. (2002) Isolation of nucleic acids from lichens. In *Protocols in Lichenology. Culturing, Biochemistry, Ecophysiology and Use in Biomonitoring* (I. Kranner, R. Beckett & A. Varma, eds): 381–391. Berlin, Heidelberg: Springer-Verlag.

Dal Grande, F., Sharma, R., Meiser, A., Rolshausen, G., Büdel, B., Mishra, B., Thines, M., Otte, J., Pfenninger, M. & Schmitt, I. (2017) Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evolutionary Biology* **17:** 93.

Dal Grande, F., Rolshausen, G., Divakar, P. K., Crespo, A., Otte, J., Schleuning, M. & Schmitt, I. (2018) Environment and host identity structure communities of green algal symbionts in lichens. *New Phytologist* **217:** 277–289.

Davydov, E. A., Peršoh, D. & Rambold, G. (2010) The systematic position of *Lasallia caroliniana* (Tuck.) Davydov, Peršoh & Rambold comb. nova and considerations on the generic concept of *Lasallia* (*Umbilicariaceae*, Ascomycota). *Mycological Progress* **9:** 261–266.

Debuchy, R. & Turgeon, B. G. (2006) Mating-type structure, evolution, and function in Euascomycetes. In *The Mycota I. Growth, Differentiation and Sexuality* (U. Kües & R. Fischer, eds): 293–323. Berlin, Heidelberg: Springer-Verlag.

Denison, W. C. (2003) Apothecia and ascospores of *Lobaria oregana* and *Lobaria pulmonaria* investigated. *Mycologia* **95:** 513–518.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K.-S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L. & Metcalf, W. W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology* **10:** 963–968.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2:** 953–971.

Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. (2011) Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE* **6:** e16526.

Greshake Tsovaras, B. (2018) *Characterizing the hologenome of* Lasallia pustulata *and tracing genomic footprints of lichenization.* Ph.D. thesis, Goethe Universität Frankfurt.

Greshake Tsovaras, B., Zehr, S., Dal Grande, F., Meiser, A., Schmitt, I. & Ebersberger, I. (2016) Potential and pitfalls of eukaryotic metagenome skimming: a test case for lichens. *Molecular Ecology Resources* **16:** 511–523.

Grewe, F., Huang, J.-P., Leavitt, S. D. & Lumbsch, H. T. (2017) Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. *Scientific Reports* **7:** 9884.

Grube, M., Berg, G., Andrésson, Ó. S., Vilhelmsson, O., Dyer, P. S. & Miao, V. P. W. (2014) Lichen genomics: prospects and progress. In *The Ecological Genomics of Fungi* (F. Martin, ed.): 191–212. Hoboken, New Jersey: John Wiley & Sons.

Hestmark, G. (1992) Sex, size, competition and escape: strategies of reproduction and dispersal in *Lasallia pustulata* (*Umbilicariaceae*, Ascomycetes). *Oecologia* **92:** 305–312.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V. & Quesneville, H. (2014) PASTEC: an automatic transposable element classification tool. *PLoS ONE* **9:** e91929.

Honegger, R., Zippler, U., Gansner, H. & Scherrer, S. (2004) Mating systems in the genus *Xanthoria* (lichen-forming ascomycetes). *Mycological Research* **108:** 480–488.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. & Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110:** 462–467.

Kampa, A., Gagunashvili, A. N., Gulder, T. A. M., Morinaka, B. I., Daolio, C., Godejohann, M., Miao, V. P. W., Piel, J. & Andresson, O. S. (2013) Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proceedings of the National Academy of Sciences of the United States of America* **110:** E3129–E3137.

Krijger, J.-J., Thon, M. R., Deising, H. B. & Wirsel, S. G. (2014) Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. *BMC Genomics* **15:** 722.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305:** 567–580.

Kronstad, J. W. & Staben, C. (1997) Mating type in filamentous fungi. *Microbiology* **31:** 245–276.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004) Versatile and open software for comparing large genomes. *Genome Biology* **5:** R12.

Leavitt, S. D., Grewe, F., Widhelm, T., Muggia, L., Wray, B. & Lumbsch, H. T. (2016) Resolving evolutionary relationships in lichen-forming fungi using diverse phylogenomic datasets and analytical approaches. *Scientific Reports* **6:** 22262.

Ludwig, L. R., Summerfield, T. C., Lord, J. M. & Singh, G. (2017) Characterization of the mating-type locus (*MAT*) reveals a heterothallic mating system in *Knightiella splachnirima*. *Lichenologist* **49:** 373–385.

Magain, N., Miadlikowska, J., Mueller, O., Gajdeczka, M., Truong, C., Salamov, A. A., Dubchak, I., Grigoriev, I. V., Goffinet, B., Sérusiaux, E., *et al.* (2017) Conserved genomic collinearity as a source of broadly applicable, fast evolving, markers to resolve species complexes: a case study using the lichen-forming genus *Peltigera* section *Polydactylon*. *Molecular Phylogenetics and Evolution* **117:** 10–29.

Meiser, A., Otte, J., Schmitt, I. & Dal Grande, F. (2017) Sequencing genomes from mixed DNA samples – evaluating the metagenome skimming approach in lichenized fungi. *Scientific Reports* **7:** 14881.

Moreno-Hagelsieb, G. & Latimer, K. (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24:** 319–324.

Nilsson, M. A. (2016) The devil is in the details: transposable element analysis of the Tasmanian devil genome. *Mobile Genetic Elements* **6:** e1119926.

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M. M., Gormley, N. A. & Cox, A. J. (2015)

NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* **31:** 2035–2037.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8:** 785–786.

Posner, B., Feige, G. B. & Leuckert, C. (1991) Beiträge zur chemie der flechtengattung *Lasallia* Merat. *Zeitschrift für Naturforschung C* **46:** 19–27.

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M. & Anxolabehere, D. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Computational Biology* **1:** 166–175.

Rolshausen, G., Dal Grande, F., Sadowska-Deś, A. D., Otte, J. & Schmitt, I. (2018) Quantifying the climatic niche of symbiont partners in a lichen symbiosis indicates mutualist-mediated niche expansions. *Ecography* DOI: 10.1111/ecog.03457

Sancho, L. G. & Crespo, A. (1989) *Lasallia hispanica* and related species. *Lichenologist* **21:** 45–58.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210–3212.

Singh, G., Dal Grande, F., Cornejo, C., Schmitt, I. & Scheidegger, C. (2012) Genetic basis of self-incompatibility in the lichen-forming fungus *Lobaria pulmonaria* and skewed frequency distribution of mating-type idiomorphs: implications for conservation. *PLoS ONE* **7:** e51402.

Singh, G., Dal Grande, F., Werth, S. & Scheidegger, C. (2015) Long-term consequences of disturbances on reproductive strategies of the rare epiphytic lichen *Lobaria pulmonaria*: clonality a gift and a curse. *FEMS Microbiology Ecology* **91:** 1–11.

Soderlund, C., Bomhoff, M. & Nelson, W. M. (2011) SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Research* **39:** e68.

Steinhäuser, S. S., Andrésson, Ó. S., Pálsson, A. & Werth, S. (2016) Fungal and cyanobacterial gene expression in a lichen symbiosis: effect of temperature and location. *Fungal Biology* **120:** 1194–1208.

Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in Microbiology* **3:** 410.

Vivas, M., Pérez-Ortega, S., Pintado, A. & Sancho, L. G. (2017) F_v/F_m acclimation to the Mediterranean summer drought in two sympatric *Lasallia* species from the Iberian mountains. *Lichenologist* **49:** 157–165.

Wang, Y.-Y., Liu, B., Zhang, X.-Y., Zhou, Q.-M., Zhang, T., Li, H., Yu, Y.-F., Zhang, X.-L., Hao, X.-Y., Wang, M., *et al.* (2014) Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (*Verrucariales*, Ascomycota). *BMC Genomics* **15:** 34.

Ward, N. & Moreno-Hagelsieb, G. (2014) Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS ONE* **9:** 1–6.

Wessels, J. G. H. (1993) Wall growth, protein excretion and morphogenesis in fungi. *New Phytologist* **123:** 397–413.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8:** 973–982.

Wolfe, K. H. & Li, W.-H. (2003) Molecular evolution meets the genomics revolution. *Nature Genetics* **33:** 255–265.

Xavier, B. B., Miao, V. P. W., Jónsson, Z. O. & Andrésson, Ó. S. (2012) Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: features and phylogeny. *Fungal Biology* **116:** 802–814.

Yamamoto, Y., Mizuguchi, R. & Yamada, Y. (1985) Tissue cultures of *Usnea rubescens* and *Ramalina yasudae* and production of usnic acid in their cultures. *Agricultural and Biological Chemistry* **49:** 3347–3348.

Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30:** 614–620.

Supplementary material to

Dal Grande, F., Meiser, A., Greshake Tzovaras, B., Otte, J., Ebersberger, I., & Schmitt, I. (2018a). The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *The Lichenologist*, 50(3), 329–340, doi:10.1017/S002428291800021X.

Table S1. List of transposable elements (TEs) detected in the genomes of *L. pustulata* and *L. hispanica*.
Table S2. Putatively secreted proteins and their annotation in *L. hispanica*.
Table S3. Biosynthetic gene and gene clusters in *L. hispanica* and *L. pustulata*.
Table S4. Domain architecture of type I PKS genes in *L. hispanica*.
Table S5. Orthologs mapped as reciprocal best BLAST hits (RBHs) between *L. hispanica* and *L. pustulata*.[1]

[1] This supplementary material can be found online accompanying the published paper at http://dx.doi.org/10.1017/S002428291800021X.

Table S1. List of transposable elements (TEs) detected in the genomes of *L. pustulata* and *L. hispanica*. TEs are named following the classification by Wicker *et al.* (2007).

		Lasallia pustulata		*Lasallia hispanica*		
TE	**Elements**	**Copies**	**Full copies**	**Copies**	**Full copies**	**Definition**
RLX-TRIM	1	11	4	0	0	*Retrotransposon LTR TRIM*
RLX-LARD	1	25	1	0	0	*Retrotransposon LTR LARD*
RLX	16	550	46	385	0	*Retrotransposon LTR*
RLC	25	865	166	276	0	*Retrotransposon LTR Copia*
RIX	1	12	4	1	0	*Retrotransposon non-LTR LINE*
RII	4	143	29	82	0	*Retrotransposon non-LTR SINE*
DXX-MITE	2	31	7	5	0	*DNA Transposon MITE*
DXX	2	99	4	13	0	*DNA Transposon*
DTX	18	824	86	96	1	*DNA Transposon TIR*
DTT	4	139	28	15	0	*DNA Transposon TIR Mariner*
DTB	1	10	4	3	0	*DNA Transposon TIR PiggyBac*
DTA	1	80	4	16	0	*DNA Transposon TIR hAT*
DHX	7	680	23	108	0	*Helitron*
NoCat	14	1191	83	209	0	*unknown*
Total Class I	70	1606	250	744	0	
Total Class II	35	1863	156	256	1	
Total consensus	121	4660	489	1209	1	

Table S2. Putatively secreted proteins and their annotation in *L. hispanica*.

Gene	Description	Length	Similarity mean (%)
Lashi_FUN_08957	agmatinase 1	61	70.65
Lashi_FUN_05102	S1 P1 nuclease	110	70.05
Lashi_FUN_09159	methionyl-trna synthetase	136	87.05
Lashi_FUN_06809	fk506-binding 2	137	81.20
Lashi_FUN_00822	phosphatidylglycerol phosphatidylinositol transfer	166	72.45
Lashi_FUN_05836	dna repair	182	71.40
Lashi_FUN_04673	Small COPII coat GTPase SAR1	189	97.80
Lashi_FUN_05808	endosomal P24B	203	89.40
Lashi_FUN_04738	period circadian	204	68.35
Lashi_FUN_04189	secretory pathway Ssp120	205	75.30
Lashi_FUN_05518	carbohydrate esterase family 5	226	68.10
Lashi_FUN_01866	jelly roll fold-like	229	76.15
Lashi_FUN_08695	Protein of unknown function DUF4188	235	57.75
Lashi_FUN_05602	predicted protein	243	62.35
Lashi_FUN_01524	beta-1,6-glucan biosynthesis (Knh1)	257	68.85
Lashi_FUN_08492	extracellular glucoamylase	266	83.30
Lashi_FUN_03243	glycoside hydrolase family 17	292	72.20
Lashi_FUN_03439	chitinase 3	320	60.35
Lashi_FUN_00266	glycoside hydrolase family 43	322	66.50
Lashi_FUN_04931	hydroxyindole o	328	72.90
Lashi_FUN_02433	fad binding domain	345	63.80
Lashi_FUN_00084	disulfide-isomerase tigA	364	80.90
Lashi_FUN_00046	glycoside hydrolase family 16	371	64.10
Lashi_FUN_04754	Arginase deacetylase	376	83.50
Lashi_FUN_02883	glycoside hydrolase family 45	378	69.25
Lashi_FUN_00791	leucine aminopeptidase 1	388	84.50
Lashi_FUN_02678	agmatinase 1	393	90.70
Lashi_FUN_06734	gpi-anchor transamidase precursor	394	84.35
Lashi_FUN_00994	Zn-dependent exopeptidase	398	74.25
Lashi_FUN_01306	vacuolar protease A	402	89.90
Lashi_FUN_01843	chaperone dnaj	409	83.90
Lashi_FUN_04243	dna repair	410	69.80
Lashi_FUN_06366	major royal jelly	420	72.95
Lashi_FUN_07537	glycoside hydrolase family 5	422	69.35
Lashi_FUN_00389	Glucan 1,3-beta-glucosidase	423	72.60
Lashi_FUN_06838	acid protease	429	64.65
Lashi_FUN_06397	acid phosphatase	445	76.40
Lashi_FUN_08433	major royal jelly	445	74.95
Lashi_FUN_07269	alpha beta-hydrolase	446	66.25
Lashi_FUN_01614	agmatinase	454	86.20
Lashi_FUN_01702	PLC-like phosphodiesterase	454	70.65
Lashi_FUN_05140	glycoside hydrolase family 76	458	71.65
Lashi_FUN_00391	murein transglycosylase	460	74.75
Lashi_FUN_05490	GPI-anchored cell wall beta-1,3-endoglucanase	471	79.00
Lashi_FUN_04639	tpr repeat	472	84.45
Lashi_FUN_03007	carbohydrate esterase family 4	474	63.10
Lashi_FUN_04851	FAS1 domain-containing	477	72.75

– continued on next page

Table S2 – *continued from previous page*

Gene	Description	Length	Similarity mean (%)
Lashi_FUN_06262	carboxypeptidase s1	477	83.70
Lashi_FUN_06826	ATP-dependent DNA helicase II subunit 2	480	78.80
Lashi_FUN_05324	glycoside hydrolase family 16	483	74.20
Lashi_FUN_04966	SUN domain (Uth1)	484	60.35
Lashi_FUN_04297	oligosaccharyltransferase alpha subunit	487	77.90
Lashi_FUN_00924	glycoside hydrolase family 5	488	75.00
Lashi_FUN_05664	type I phosphodiesterase nucleotide pyrophosphatase	488	79.05
Lashi_FUN_01489	disulfide isomerase	490	69.60
Lashi_FUN_03662	autophagic serine protease alp2	491	84.00
Lashi_FUN_03054	Alpha-galactosidase	501	73.25
Lashi_FUN_03861	Peptidase serine carboxypeptidase	502	61.70
Lashi_FUN_02411	cytochrome P450 oxidoreductase alkane hydroxylase	510	66.05
Lashi_FUN_06626	misfolded glyco s degradation	510	57.80
Lashi_FUN_07822	Glucooligosaccharide oxidase	510	74.75
Lashi_FUN_04318	Actin patch 1	519	79.80
Lashi_FUN_02201	glycoside hydrolase family 125	531	73.55
Lashi_FUN_02379	carboxypeptidase cpds	531	80.00
Lashi_FUN_04326	probable lysophospholipase precursor	538	75.45
Lashi_FUN_05340	disulfide-isomerase domain	538	83.10
Lashi_FUN_05761	3-phytase A	540	82.45
Lashi_FUN_03287	SUN domain (Adg3)	546	74.70
Lashi_FUN_01766	phosphoglycerate mutase	548	77.25
Lashi_FUN_06834	glycoside hydrolase family 125	549	81.80
Lashi_FUN_04090	carboxylesterase family	562	72.40
Lashi_FUN_07843	carboxypeptidase S1	567	80.20
Lashi_FUN_03842	Aspartic-type endopeptidase	568	68.80
Lashi_FUN_05489	carboxypeptidase S1	575	79.75
Lashi_FUN_07655	hypothetical protein W97_08084	575	53.45
Lashi_FUN_07923	gamma-glutamyltransferase	576	78.05
Lashi_FUN_05423	glycoside hydrolase family 32	577	73.50
Lashi_FUN_04250	glycoside hydrolase family 20	585	64.75
Lashi_FUN_01895	multicopper oxidase	589	74.50
Lashi_FUN_06837	conidial pigment biosynthesis oxidase abr1 brown 1	594	70.65
Lashi_FUN_05261	Amidase	609	80.90
Lashi_FUN_08045	extracellular glucoamylase	616	78.75
Lashi_FUN_05267	versicolorin B synthase	622	68.40
Lashi_FUN_02380	tripeptidyl peptidase sed3	633	73.70
Lashi_FUN_05770	GMC oxidoreductase	637	72.45
Lashi_FUN_05514	lysophospholipase plb1	653	78.00
Lashi_FUN_07621	subtilisin	665	76.75
Lashi_FUN_04896	Metallo-dependent phosphatase	679	78.60
Lashi_FUN_00069	amidase signature enzyme	680	71.70
Lashi_FUN_01787	carboxypeptidase s1	713	72.95
Lashi_FUN_05799	calcineurin-like phosphoesterase	727	67.60
Lashi_FUN_04412	glycoside hydrolase family 55	798	76.25
Lashi_FUN_03433	mannosyl-oligosaccharide glucosidase	819	79.95
Lashi_FUN_06388	glycoside hydrolase family 55	833	76.10
Lashi_FUN_07159	glycoside hydrolase family 3	884	79.20

– *continued on next page*

Table S2 – *continued from previous page*

Gene	Description	Length	Similarity mean (%)
Lashi_FUN_05128	carbohydrate-binding module family 18	897	76.35
Lashi_FUN_07869	neutral alpha-glucosidase ab	961	79.10
Lashi_FUN_03301	m repeat	1002	73.85
Lashi_FUN_08174	acid trehalase	1053	73.45
Lashi_FUN_06754	ATP dependent dna ligase domain	1103	65.25
Lashi_FUN_07823	heme peroxidase	1241	71.85
Lashi_FUN_08543	glycoside hydrolase family 18	1296	65.40
Lashi_FUN_00665	udp-glucose:glyco glucosyltransferase	1520	75.55
Lashi_FUN_02646	WSC-domain-containing	1672	69.95

Table S3. Biosynthetic gene and gene clusters in *L. hispanica* and *L. pustulata*. Green boxes indicate core biosynthetic gene orthologs (RBH); gray boxes indicate orthologous (RBH) tailoring enzymes. Cluster names in bold-face indicate orthologous gene clusters (see Fig. 4).

	L. hispanica			*L. pustulata*		
	Cluster	**Gene ID**	**fungiSMASH annotation**	**Clusters**	**Gene ID**	**fungiSMASH annotation**
a) Non-reducing PKS gene clusters	**Cluster 3**	Lashi _FUN_01884	KS - AT - ACP	**Cluster 18**	FUN_05144	KS - AT - ACP
		Lashi _FUN_01882	acyltransferase		FUN_05146	acyltransferase
		Lashi _FUN_01886	metallo-beta-lactamase family protein		FUN_05142	metallo-beta-lactamase family protein
		Lashi _FUN_01887	halogenase		FUN_05141	halogenase
	Cluster 6	Lashi _FUN_02945	KS - AT - ACP	**Cluster 14**	FUN_04869	KS - AT - ACP
		Lashi _FUN_02939	aldo/keto reductase family oxidoreductase		FUN_04875	aldo/keto reductase family oxidoreductase
		Lashi _FUN_02940	Drug resistance transporter, EmrB/QacA		FUN_04874	Drug resistance transporter, EmrB/QacA
		Lashi _FUN_02941	cytochrome P450		FUN_04873	cytochrome P450
		Lashi _FUN_02944	metallo-beta-lactamase family protein		FUN_04870	metallo-beta-lactamase family protein
		Lashi _FUN_02947	O-methyltransferase		FUN_04866	O-methyltransferase
		Lashi _FUN_02948	haloalkane dehalogenase		FUN_04863	
					FUN_04879	transcriptional regulator, SARP family
	Cluster 9	Lashi _FUN_05414	KS - AT - ACP - ACP - TE	**Cluster 1**	FUN_00584	KS - AT - ACP - ACP - TE
		Lashi _FUN_05412	dioxygenase, TauD/TfdA		FUN_00586	
		Lashi _FUN_05413	FAD linked oxidase domain protein		FUN_00585	FAD linked oxidase domain protein
		Lashi _FUN_05417	aldo/keto reductase family oxidoreductase		FUN_00580	aldo/keto reductase family oxidoreductase
	Cluster 13	Lashi _FUN_07257	KS - AT - ACP - ACP - TE	**Cluster 2**	FUN_00778	KS - AT - ACP - ACP - TE
		Lashi _FUN_07256	serine/threonine protein kinase		FUN_00779	serine/threonine protein kinase
	Cluster 15	Lashi _FUN_03177	KS - AT - ACP - ACP - TE	Cluster 19	FUN_05257	KS - AT - ACP - ACP - TE
					FUN_05249	transporter, EamA family
					FUN_05255	cytochrome P450
				Cluster 28	FUN_09664	KS - AT - ACP - TE
					FUN_09663	short-chain dehydrogenase/reductase SDR
				Cluster 22	FUN_07319	KS - AT - ACP - ACP - cMT - TD
					FUN_07313	FAD linked oxidase domain protein
					FUN_07315	short-chain dehydrogenase/reductase SDR
					FUN_07317	major facilitator transporter
					FUN_07320	crotonyl-CoA reductase / alcohol dehydrogenase
					FUN_07322	FAD linked oxidase domain protein
					FUN_07325	Drug resistance transporter, EmrB/QacA

– continued on next page

Table S3 – *continued from previous page*

	L. hispanica			*L. pustulata*		
	Cluster	**Gene ID**	**fungiSMASH annotation**	**Cluster**	**Gene ID**	**fungiSMASH annotation**
b) Reducing PKS gene clusters	**Cluster 4**	Lashi _FUN_07807	KS - AT - DH - cMT - ER - KR	**Cluster 21**	FUN_05480	KS - AT - DH - cMT - ER - KR
		Lashi _FUN_07800	short-chain dehydrogenase/reductase SDR		FUN_05471	
		Lashi _FUN_07802	acyl-CoA dehydrogenase		FUN_05475	acyl-CoA dehydrogenase
		Lashi _FUN_07803	unannotated orf		FUN_05476	
		Lashi _FUN_07804	AMP-dependent synthetase and ligase		FUN_05477	AMP-dependent synthetase and ligase
		Lashi _FUN_07808	FAD linked oxidase domain protein		FUN_05489	FAD dependent oxidoreductase
					FUN_05481	FAD linked oxidase domain protein
					FUN_05483	Drug resistance transporter, EmrB/QacA
	Cluster 10	Lashi _FUN_06857	KS - AT - DH - ACP - TE	**Cluster 23**	FUN_07066	KS - AT - DH - ACP - TE
		Lashi _FUN_06858	O-methyltransferase		FUN_07067	O-methyltransferase
	Cluster 14	Lashi _FUN_07789	KS - AT - DH - KR - ACP	Cluster 11	FUN_02486	KS - AT - DH - KR - ACP
		Lashi _FUN_07784	major facilitator transporter		FUN_02488	phenylalanine-specific permease
		Lashi _FUN_07788	cytochrome P450		FUN_02493	serine/threonine protein kinase
	Cluster 1	Lashi _FUN_05771	KS - AT - DH - ACP - cMT	Cluster 12	FUN_04144	KS - AT - DH - ACP - ACP - cMT
		Lashi _FUN_05765	O-methyltransferase		FUN_04142	cytochrome P450
		Lashi _FUN_05766	short-chain dehydrogenase/reductase SDR		FUN_04143	monooxygenase FAD-binding
		Lashi _FUN_05767	Drug resistance transporter, EmrB/QacA		FUN_04145	AMP-dependent synthetase and ligase
		Lashi _FUN_05770	pyridine nucleotide-disulfide oxidoreductase		FUN_04146	O-methyltransferase
		Lashi _FUN_05772	AMP-dependent synthetase and ligase		FUN_04148	short-chain dehydrogenase/reductase SDR
		Lashi _FUN_05773	cytochrome P450			
				Cluster 3	FUN_00874	KS - AT - DH - ER - KR - ACP
	Cluster 2	Lashi _FUN_07360	KS - AT - DH - KR - ACP		FUN_00866	cytochrome P450
		Lashi _FUN_07362	beta-lactamase		FUN_00869	cytochrome P450
		Lashi _FUN_07363	sensor histidine kinase		FUN_00871	Beta-ketoacyl synthase (redPKS with missing ACP)
					FUN_00872	Drug resistance transporter, EmrB/QacA
	Cluster 7	Lashi _FUN_08583	KS - AT - DH - cMT - ER - KR - ACP		FUN_00878	metallo-beta-lactamase family protein
		Lashi _FUN_08584	Drug resistance transporter, EmrB/QacA			
				Cluster 4	FUN_01125	KS - AT - DH - ACP - ACP - TE
c) Non-reducing / reducing PKS gene clusters	Cluster 5	Lashi _FUN_08273	KS - AT - ACP - TE			
		Lashi _FUN_08274	KS - AT - DH - ER - KR - ACP			
		Lashi _FUN_08275	Drug resistance transporter, EmrB/QacA			
	Cluster 8	Lashi _FUN_08739	KS - AT - ACP - ACP - TE			
		Lashi _FUN_08740	KS - AT - DH - ER - KR - ACP			

– *continued on next page*

Table S3 – *continued from previous page*

	L. hispanica			*L. pustulata*		
	Cluster	**Gene ID**	**fungiSMASH annotation**	**Cluster**	**Gene ID**	**fungiSMASH annotation**
d) Terpenes	**Cluster 17**	Lashi _FUN_07446	terpene	**Cluster 24**	FUN_08059	terpene
		Lashi _FUN_07447	phytoene dehydrogenase		FUN_08060	phytoene dehydrogenase
	Cluster 18	Lashi _FUN_02410	terpene	**Cluster 8**	FUN_01960	terpene
		Lashi _FUN_02411	cytochrome P450 oxidoreductase alkane hydroxylase		FUN_01959	cytochrome P450 oxidoreductase alkane hydroxylase
	Cluster 19	Lashi _FUN_04980	terpene	**Cluster 25**	FUN_08457	terpene
	Cluster 20	Lashi _FUN_06819	terpene	**Cluster 17**	FUN_03791	terpene
		Lashi _FUN_06818	peptidase s9b dipeptidylpeptidase iv domain		FUN_03792	peptidase s9b dipeptidylpeptidase iv domain
				Cluster 15	FUN_06007	
e) Lantipeptide gene cluster				Cluster 6	FUN_01532	lanthionine synthetase C family protein
					FUN_01529	phosphoglycerate mutase
					FUN_01535	cytochrome P450
f) Type III PKS gene cluster	**Cluster 21**	Lashi _FUN_04513	chalcone and stilbene synthase domain protein	**Cluster 16**	FUN_06094	chalcone and stilbene synthase domain protein
		Lashi _FUN_04514	cytochrome P450		FUN_06095	cytochrome P450
g) Type I PKS - NRPS hybrid gene clusters				Cluster 26	FUN_09083	KS - AT - DH - cMT - KR - C - AMP - PCP - TD
					FUN_09079	Drug resistance transporter, EmrB/QacA
					FUN_09080	crotonyl-CoA reductase / alcohol dehydrogenase
					FUN_09081	cytochrome P450
h) Partial reducing type I PKS gene clusters	Cluster 11	Lashi _FUN_07207	KS - AT - DH - cMT	Cluster 20	FUN_05310	sugar transport protein
		Lashi _FUN_07206	short-chain dehydrogenase/reductase SDR		FUN_05314	KS - AT - DH - cMT - ER - KR
		Lashi _FUN_07208	8-amino-7-oxononanoate synthase		FUN_05316	putative carboxymuconolactone decarboxylase
					FUN_05321	Drug resistance transporter, EmrB/QacA
	Cluster 12	Lashi _FUN_07044	KS - AT - DH - ER - KR			
		Lashi _FUN_07042	short-chain dehydrogenase/reductase SDR	Cluster 27	FUN_09129	KS - AT - DH
		Lashi _FUN_07043	Drug resistance transporter, EmrB/QacA		FUN_09130	ER - KR - ACP (crotonyl-CoA reductase / alcohol dehydrogenase)
					FUN_09128	KS
	Cluster 16	Lashi _FUN_04483	KS - AT		FUN_09125	short-chain dehydrogenase/reductase SD
		Lashi _FUN_04479				
		Lashi _FUN_04481	putative carboxymuconolactone decarboxylase	Cluster 3	FUN_02125	KS - AT - DH
					FUN_02124	cMT - ER - KR - ACP

KS: Keto-synthase; AT: Acyltransferase; ACP: Acyl carrier protein; TE: Thioesterase; DH: Dehydratase; cMT: C-Methyltransferase; ER: Enoylreductase; KR: Ketoreductase

Table S4. Domain architecture of type I PKS genes in *L. hispanica*.

Non-reducing PKS genes	Domain annotation
Lashi_FUN_01884	KS - AT - ACP
Lashi_FUN_02945	KS - AT - ACP
Lashi_FUN_05414	KS - AT - ACP - ACP - TE
Lashi_FUN_07257	KS - AT - ACP - ACP - TE
Lashi_FUN_03177	KS - AT - ACP - ACP - TE

Reducing PKS genes	Domain annotation
Lashi_FUN_05771	KS - AT - DH - ACP - cMT
Lashi_FUN_06857	KS - AT - DH - ACP - TE
Lashi_FUN_07807	KS - AT - DH - cMT - ER - KR
Lashi_FUN_08583	KS - AT - DH - cMT - ER - KR - ACP
Lashi_FUN_07789	KS - AT - DH - KR - ACP
Lashi_FUN_07360	KS - AT - DH - KR - ACP

KS: Keto-synthase; AT: Acyltransferase; ACP: Acyl carrier protein; TE: Thioesterase; DH: Dehydratase; cMT: C-Methyltransferase; ER: Enoylreductase; KR: Ketoreductase

A.3. Biosynthetic gene clusters in two metabolite-rich species

Statement of author contributions

Title: Biosynthetic gene content of the 'perfume lichens' *Evernia prunastri* and *Pseudevernia furfuracea*

Journal: Molecules 24(1), 203, doi:10.3390/molecules24010203

Status: Accepted on 4 January 2019

Authors: Anjuli Calchera (AC), Francesco Dal Grande (FDG), Helge B. Bode (HBB), Imke Schmitt (IS)

Author contributions:

(1) Development and planning	
Doctoral candidate (AC):	60 %
Coauthor FDG:	20 %
Coauthor IS:	20 %

(2) Conduction of experiments
n/a

(3) Compilation of data sets and figures/tables	
Doctoral candidate (AC):	100 % Compilations of data sets (genomes, reference database), preparation of all figures and tables

(4) Data analyses and interpretation	
Doctoral candidate (AC):	80 % All data analysis (gene prediction, annotation, genome mining, comparative mapping, phylogenetic analysis), data interpretation
Coauthor FDG:	5 % Help with phylogenetic analyses, data interpretation
Coauthor IS:	10 % Data interpretation
Coauthor HBB:	5 % Data interpretation

(5) Preparation of manuscript	
Doctoral candidate (AC):	90 %
Coauthor FDG:	5 %
Coauthor IS:	5 %

molecules

MDPI

Article

Biosynthetic Gene Content of the 'Perfume Lichens' *Evernia prunastri* and *Pseudevernia furfuracea*

Anjuli Calchera [1,2], Francesco Dal Grande [2], Helge B. Bode [3] and Imke Schmitt [1,2,*]

1 Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; Anjuli.Calchera@senckenberg.de
2 Senckenberg Biodiversity and Climate Research Centre (S-BiKF), Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt am Main, Germany; Francesco.DalGrande@senckenberg.de
3 Molecular Biotechnology, Department for Biosciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; h.bode@bio.uni-frankfurt.de
* Correspondence: Imke.Schmitt@senckenberg.de; Tel.: +49-(0)69-7542-1855

Received: 14 December 2018; Accepted: 4 January 2019; Published: 8 January 2019

check for updates

Abstract: Lichen-forming fungi produce a vast number of unique natural products with a wide variety of biological activities and human uses. Although lichens have remarkable potential in natural product research and industry, the molecular mechanisms underlying the biosynthesis of lichen metabolites are poorly understood. Here we use genome mining and comparative genomics to assess biosynthetic gene clusters and their putative regulators in the genomes of two lichen-forming fungi, which have substantial commercial value in the perfume industry, *Evernia prunastri* and *Pseudevernia furfuracea*. We report a total of 80 biosynthetic gene clusters (polyketide synthases (PKS), non-ribosomal peptide synthetases and terpene synthases) in *E. prunastri* and 51 in *P. furfuracea*. We present an in-depth comparison of 11 clusters, which show high homology between the two species. A ketosynthase (KS) phylogeny shows that biosynthetic gene clusters from *E. prunastri* and *P. furfuracea* are widespread across the Fungi. The phylogeny includes 15 genomes of lichenized fungi and all fungal PKSs with known functions from the MIBiG database. Phylogenetically closely related KS domains predict not only similar PKS architecture but also similar cluster architecture. Our study highlights the untapped biosynthetic richness of lichen-forming fungi, provides new insights into lichen biosynthetic pathways and facilitates heterologous expression of lichen biosynthetic gene clusters.

Keywords: lichen secondary metabolites; tree moss; oakmoss; biosynthetic gene clusters; polyketide synthases; non-ribosomal peptide synthetases; terpene synthases; transcription factor; phylogeny; comparative genomics

1. Introduction

Lichens are symbioses, composed of a fungal partner (mycobiont) and one or more photosynthetic partners (photobiont) [1]. Approximately one fifth of all described fungi form lichens and more than 19,000 lichen species are described [2]. Lichens can be found in most terrestrial ecosystems. They impact various community processes, also because of their rich and diverse secondary chemistry [3,4]. Recent studies have shown that lichens represent complex multi-species symbiotic assemblages, forming microhabitats and harbouring a high diversity of other eukaryotic and prokaryotic microorganisms [5–7]. High-throughput sequencing technologies have revealed the presence of multiple fungal and algal species, along with hyper-diverse bacterial communities, within individual lichen thalli [7–13].

Lichens produce a great variety of natural compounds, and more than 1000 secondary metabolites have been identified to date, many of which are exclusively found in lichens [14–17]. The vast

majority of these characteristic lichen secondary metabolites is of fungal origin. Many lichen secondary metabolites have important ecological roles including light-screening, chemical weathering, allelopathic and anti-herbivore defence [18–20]. Furthermore, lichens are a promising source for pharmaceutically interesting natural products, because of the manifold biological activities of lichen compounds—including antiviral, antibiotic, antitumor, allergenic, plant growth inhibitory, antiherbivore and enzyme inhibitory activities [15,19,21,22]. For example, the secondary metabolites atranorin, evernic acid, physodic acid and usnic acid found in the lichens *Evernia prunastri* and *Pseudevernia prunastri* are strong metabolic enzyme inhibitors, and atranorin may inhibit lung cancer cell mobility and tumorigenesis [22,23]. A wide range of lichens have been used in traditional medicines all around the world [24]. Today, one of the more significant economic uses of lichens is in the perfume industry. Combined lichen material is used for extracts added to perfumes for a sweet and mossy smell and to ensure persistence on the skin. Several hundred to thousands of tons of *E. prunastri* and *P. furfuracea* are harvested every year in France, Morocco and South-eastern Europe for this purpose [16,19,25,26].

The most abundant classes of lichen secondary metabolites are phenolic compounds, dibenzofurans, depsides and depsidones built by multienzyme polyketide synthases (PKS). Derived monoaromatic subunits (orcinol, β-orcinol type or methylphloroacetophenone) are then linked by ester, ether or carbon-carbon bonds [1,22,27]. Fungal PKSs consist of a set of active site domains used in an iterative fashion for multiple catalytic cycles and are subdivided into non-reducing (NR-PKS), or reducing (R-PKS) according to the level of reductive processing [28]. Both types of polyketides are found in lichen-forming fungi but fully oxidized polyketides account for most of the well-known and common lichen substance classes such as depsides and depsidones whose formation is controlled by NR-PKS genes [16,29]. Fungal secondary metabolic pathway genes are often grouped in gene clusters including genes encoding tailoring functions, transporters, and pathway-specific regulatory genes along with the core genes [30–32].

Recent advances in genome sequencing and bioinformatic mining algorithms have enabled the identification and characterization of biosynthetic gene clusters from genome sequences and led to the discovery of thousands of biosynthetic gene clusters [33–36]. However, the identity, structure and function of the metabolites encoded by these gene clusters remain mostly unknown. The clusters are therefore referred to as cryptic or orphan clusters [37]. In particular, fungal PKS gene clusters are challenging to functionally characterize, mainly due to the unpredictable intrinsic cryptic programming concealed within iterative PKS genes themselves and missing knowledge on precise selectivity and activity of tailoring enzymes [30,31,38]. From the vast biosynthetic potential found in genomes only a small fraction has been investigated experimentally to connect genes to metabolite production—for example through gene knockout or heterologous expression. But the presence of cryptic or orphan gene clusters in many genomes, and the fact that a vast number of biosynthetic genes is not reflected by the metabolic profile, hint at a plethora of yet undiscovered chemical compounds [32,39,40].

Despite the rich chemical diversity found in lichens, only few biosynthetic studies are available. In lichen-forming fungi the slow growth rates and difficulties in aposymbiotic mycobiont cultivation further hinder the characterization of the molecular mechanisms underlying the biosynthesis of characteristic lichen metabolites [14,41]. The biosynthesis of substances may depend on abiotic factors such as geographic, altitudinal or microhabitat conditions or the response to microclimatic fluctuations, seasonality, chemical signals, hydration, or UV radiation [14,22]. Biosynthesis of natural products may also be influenced by biological factors, such as the presence of competing plant or lichen species, predation by insects, or contact to other symbionts that are part of the lichen [27]. Furthermore, lichen mycobionts synthesize significant quantities of secondary metabolites only under permissive conditions and therefore the production in axenic cultures can differ substantially from that in nature [21,22,42].

In lichen-forming fungi no secondary metabolite has been directly linked with experimental proof (i.e., gene knockout, heterologous expression) to a biosynthetic gene cluster, although several assignments

have been proposed. The first identification of a lichen secondary metabolic gene cluster—grayanic acid in *Cladonia grayi*—was accomplished by Armaleo et al. [43] combining phylogenetic analysis, biosynthetic prediction from PKS domain arrangement and correlation of mRNA levels with metabolic production. Later, the putative gene clusters encoding usnic acid and 6-hydroxymellein in *Cladonia uncials* were identified by Abdel-Hameed et al. [41,44] on the basis of PKS domain arrangement and prediction of the functions of tailoring enzymes encoded in the same gene cluster.

Several methods are applied and combined to discover PKS genes in lichenized fungi: amplification of PKS gene fragments from genomic DNA using degenerate primers and design of probes for library screening; the use of cDNA-based templates generated from reverse-transcribed mRNA, or gaining biosynthetic insights directly from transcriptomes or genomes sequenced from mycobiont cultures or reconstructed from metagenomic thalli [45–47]. The latter became possible through the advances of sequencing technologies and taxonomic assignment tools and led to the publication of several draft genomes of lichen-forming fungi over the last few years (see below). For the prediction of biosynthetic function of PKS genes there are diverse strategies, apart from experimental evidence through gene knockout, heterologous expression, radiolabelling or oxidation experiments [48–51]. These include phylogenetic approaches to infer the functions through a close phylogenetic relatedness to characterized genes [52,53], transcriptional profiling [43,47], or comparative homology mapping of entire gene clusters [31,36,54]. These approaches are often combined to provide putative assignments [47]. A review on reported PKS genes from lichenized fungi including an overview on approaches for PKS identification and characterization has been published recently and highlights the prospects of genomics-driven natural product discovery in lichens [45].

Here, we combine genome mining with comparative genetic mapping and phylogenetics of the two lichenized fungi *Evernia prunastri* and *Pseudevernia furfuracea*. Both lichens are of economic importance [1,19] and their fungal genomes contain a high number of biosynthetic gene clusters [46]. Moreover, a comprehensive review on the multitude of extracts identified from both lichens is available [25,26] and includes, for example, the bioactive metabolite atranorin detected in both lichens. The secondary metabolite richness and partially overlapping chemical profiles make these species particularly interesting study systems for combining genome mining and comparative genomics to investigate biosynthetic enzymes that may be responsible for the production of characteristic lichen substance classes. We further compare the biosynthetic gene richness found in these species with that of 13 other lichenized fungi and 57 representative fungal species and thereby provide the most complete comparison of biosynthetic gene families in lichen genomes based on entire genomes. We present mapped biosynthetic gene clusters of *E. prunastri* and *P. furfuracea* with functionally annotated accessory genes and putative regulatory genes. These carefully annotated clusters provide baseline information for heterologous expression of lichen biosynthetic gene clusters. Furthermore, we present all PKS genes mined from 15 genomes of lichenized fungi in a phylogenetic framework of characterized biosynthetic genes. Information on phylogenetic relatedness to previously characterized genes broadens our understanding of lichen biosynthetic pathways and may help us to identify promising clusters for the production of characteristic substance classes.

Specifically, we address the following questions:

(I) What is the diversity of biosynthetic gene clusters in *Evernia prunastri* and *Pseudevernia furfuracea* and how does it compare to other lichenized fungi and non-lichenized fungi?

(II) What is the architecture and gene content of those clusters with high homology between *E. prunastri* and *P. furfuracea*?

(III) Where do PKSs from *E. prunastri* and *P. furfuracea* group phylogenetically in a phylogeny of PKSs with known functions?

2. Results & Discussion

2.1. Biosynthetic Gene Richness in Fifteen Annotated Genomes of Lichen-Forming Fungi

We investigated the biosynthetic gene richness in a total of fifteen genomes of lichen-forming fungi. The species belong to different phylogenetic groups and synthesize a variety of lichen substances (Table 1).

Table 1. Overview of genomes of lichen-forming fungi used in this study.

Species	Taxonomic Group	Data Repository [1]	Gene Set Previously Available	Genome Size	Number of Scaffolds	Scaffold N50	Number of Genes	Abbreviation	Metabolites Reported [55–59]
Cladonia grayi	Lecanoromycetes, Lecanorales	JGI Clagr3 v2.0	yes	34.6 Mb	414	243,412	11,389	CGRA	grayanic acid, fumarprotocetraric acid complex
Cladonia macilenta	Lecanoromycetes, Lecanorales	NCBI AUPP00000000.1	no	37.1 Mb	240	1,469,036	10,559	CMAC	thamnolic acid, barbatic acid, didymic acid, squamatic acid, usnic acid, rhodocladonic acid
Cladonia metacorallifera	Lecanoromycetes, Lecanorales	NCBI AXCT00000000.2	no	36.7 Mb	30	1,591,850	10,497	CMET	usnic acid, didymic acid, squamatic acid, rhodocladonic acid
Cladonia uncialis	Lecanoromycetes, Lecanorales	NCBI NAPT00000000.1	no	32.9 Mb	2124	34,871	10,902	CUNC	usnic acid, squamatic acid
Endocarpon pusillum (Park et al.) [60]	Eurotiomycetes, Verrucariales	NCBI JFDM00000000.1	no	37.2 Mb	40	1,340,794	11,756	EPUP	(none reported)
Endocarpon pusillum (Wang et al.) [61]	Eurotiomycetes, Verrucariales	NCBI APWS00000000.1	yes	37.1 Mb	908	178,225	9238	EPUW	(none reported)
Evernia prunastri	Lecanoromycetes, Lecanorales	NCBI NKYR00000000.1	yes	40.3 Mb	277	264,454	10,992	EPRU	evernic acid, atranorin, usnic acid
Gyalolechia flavorubescens	Lecanoromycetes, Teloschistales	NCBI AUPK00000000.1	no	34.5 Mb	36	1,693,300	10,460	GFLA	parietin, emodin, fallacinal, fragilin
Lobaria pulmonaria	Lecanoromycetes, Peltigerales	JGI Lobpul1 v1.0	yes	56.1 Mb	1911	50,541	15,607	LPUL	stictic acid, norstictic acid, constictic acid
Pseudevernia furfuracea	Lecanoromycetes, Lecanorales	NCBI NKYQ00000000.1	yes	37.8 Mb	46	1,178,799	8842	PFUR	atranorin, olivetoric acid, physodic acid
Ramalina peruviana	Lecanoromycetes, Lecanorales	NCBI MSTJ00000000.1	no	27.0 Mb	1657	40,431	9338	RPER	sekikaic acid complex
Umbilicaria hispanica	Lecanoromycetes, Umbilicariales	NCBI PKMA00000000.1	yes	41.2 Mb	1619	145,035	8488	LHIS	gyrophoric acid, lecanoric acid, umbilicaric acid, skyrin
Umbilicaria muehlenbergii	Lecanoromycetes, Umbilicariales	NCBI JFDN00000000.1	no	34.8 Mb	7	7,009,248	8968	UMUE	gyrophoric acid
Umbilicaria pustulata	Lecanoromycetes, Umbilicariales	NCBI FWEW00000000.1	yes	39.2 Mb	3891	104,297	8268	LPUS	gyrophoric acid, lecanoric acid, hiascinic acid, skyrin
Xanthoria parietina	Lecanoromycetes, Teloschistales	JGI Xanpa2 v1.1	yes	31.9 Mb	39	1,731,186	11,065	XPAR	physcion, parietinic acid, teloschistin, emodin

[1] NCBI: National Center for Biotechnology Information; JGI: DOE Joint Genome Institute.

The genomes differ substantially in assembly status—from 7 to 3891 scaffolds—but most are 90–96% complete according to BUSCO (Figure 1a). Only the genomes of *Ramalina peruviana* (86%) and *Cladonia uncialis* (89%) score lower in completeness and also have the lowest scaffold N50 (see genome statistics given in Table 1 and Table S1). Since seven genomes did not have available gene sets, we predicted gene models. The limitations of gene prediction without transcriptome data are evident in the assessment of gene set completeness in Figure 1b. The gene sets for which we performed gene annotation are less complete (72–80% compared to 82–97%) and show more fragmented BUSCO marker genes. Nevertheless, the predicted number of genes is in the range of most of the gene sets available for lichen-forming fungi (~8200 to 11,800). The only exception here is the genome of *Lobaria pulmonaria* which has a considerable larger genome size with 56.1 Mb and 15,607 genes but also the highest rate of duplicated marker genes (Table 1, Table S1). Even though the gene prediction done here is not optimal, it improves the downstream biosynthetic gene cluster detection over the less specific annotation that can be done within antiSMASH [62].

Figure 1. (**a**) Genome completeness for assemblies and (**b**) gene set completeness for gene annotations of lichen-forming fungi. The completeness is compared against 1315 orthologous BUSCO marker genes for Ascomycota. Exact percentages can be found together with basic genome statistics in Supplementary Table S1. Asterisk indicates genomes annotated in the present study (without RNA evidence).

We identified a high number of biosynthetic gene clusters in all genomes of lichen-forming fungi (Table 2). The fungal version of antiSMASH annotated on average ~47 clusters per genome, ranging from 27 clusters for *Umbilicaria pustulata* up to as many as 80 gene clusters in *Evernia prunastri*. The most abundant family of secondary metabolite enzymes identified in all genomes are reducing type I polyketides (R-PKS) followed by terpene synthases and non-reducing type I polyketides (NR-PKS). We also detected non-ribosomal peptide synthetases (NRPS), hybrid PKS-NRPS and type III PKSs in most genomes.

Polyketides represent the most abundant class of lichen secondary metabolites and can be reduced or fully oxidized [16,22,27]. It has been reported that most polyketides found in lichens are fully oxidized [63]. Nevertheless, we find that the number of R-PKS genes exceeds by far the number of NR-PKS genes. The genomes of the lichen-forming fungi show—through their large number of secondary metabolic genes and gene clusters—the potential for a much greater number of natural products than have been reported to occur in the respective lichen species (Table 1).

Table 2. Number of biosynthetic gene clusters and main families of secondary metabolite genes found in the genomes of lichen-forming fungi.

Species	Abbreviation	Number of Clusters	Type I NR-PKS	Type I R-PKS	Type I PKS	Type III PKS	Hybrid PKS-NRPS	NRPS	Terpene Synthases	Total KS Sequences for Phylogeny	Complete PKS (KS + AT + ACP)
Cladonia grayi	CGRA	51	8	17	1	3	1	2	5	27	15
Cladonia macilenta	CMAC	52	11	23	-	2	4	2	5	38	25
Cladonia metacorallifera	CMET	59	13	26	1	2	2	2	8	42	29
Cladonia uncialis	CUNC	59	10	25	1	2	3	1	8	39	25
Endocarpon pusillum [8] (Park et al.)	EPUP	31	4	9	-	1	2	2	6	15	9
Endocarpon pusillum [9] (Wang et al.)	EPUW	31	5	12	-	1	2	1	6	19	12
Evernia prunastri	EPRU	80	9	29	1	2	4	4	13	43	30
Gyalolechia flavorubescens	GFLA	37	7	12	-	1	1	3	6	20	10
Lobaria pulmonaria	LPUL	77	8	28	-	-	4	9	16	40	22
Pseudevernia furfuracea	PFUR	51	5	23	-	2	3	4	5	31	17
Ramalina peruviana	RPER	47	9	18	3	1	1	3	2	31	15
Umbilicaria hispanica	LHIS	28	7	10	1	1	-	-	6	18	14
Umbilicaria muehlenbergii	UMUE	31	5	15	-	1	-	-	4	20	17
Umbilicaria pustulata	LPUS	27	6	9	-	1	1	-	6	16	13
Xanthoria parietina	XPAR	51	5	18	1	1	2	5	7	26	13
Sum		712	112	274	9	25	30	38	125	425	266

NR-PKS = Non-reducing type I polyketide synthase; R-PKS = Reducing type I polyketide synthase; NRPS = Non-ribosomal peptide synthetase. KS = Ketosynthase; AT = Acyltransferase; ACP = Acyl carrier protein.

The genomes of lichen-forming fungi in the Lecanoromycetes show a remarkable richness of secondary metabolic gene clusters in comparison with genomes of species from all major fungal classes. We present the number of predicted secondary metabolite gene clusters in 15 genomes of lichenized fungi (this study) and 57 ascomycete and basidiomycete genomes of non-lichenized fungi analysed in a previous study [32] (Figure 2). The average number of predicted secondary gene clusters is ~31 clusters and the average number of predicted PKS clusters is ~11. Except for two *Umbilicaria* species the number of lichen gene clusters is above average and all lichen genomes harbour more than the average amount of PKS gene cluster (~24 clusters on overage). Indeed, the highest number of PKS gene clusters are found in *Evernia prunastri* (38 PKS clusters) together with six other lichens including *Pseudevernia furfuracea* in the top ten PKS cluster count. We used a non-representative selection of fungal genomes for the comparison and thus did not include all fungal species that are reported to be rich in secondary metabolite gene clusters, for example, *Pestalotiopsis fici* (Sordariomycetes) [64]. The comparison is meant for the sole purpose of placing the richness of secondary metabolite gene clusters found in lichen-forming fungi in a broader context (Figure 2).

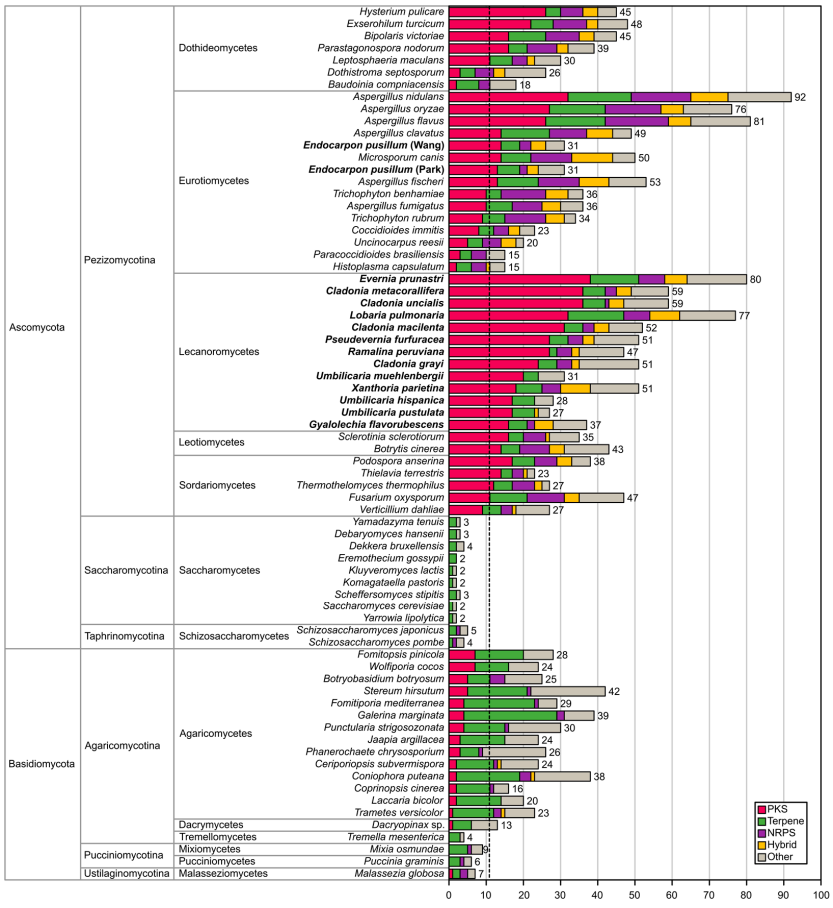


Figure 2. Overview of predicted secondary metabolic gene clusters across the genomes of representative fungal species adopted from [32]. Bold font indicates genomes of lichen-forming fungi included from this study. The dashed line shows the average number of PKS gene clusters found in a genome. "Hybrid" refers to clusters with multiple core genes belonging to different secondary metabolite families. PKS = polyketide synthase; NRPS = non-ribosomal peptide synthetase.

2.2. Gene Cluster Comparison

Pseudevernia furfuracea and especially *E. prunastri* show a high number of biosynthetic gene clusters (Table 2), and a high number of natural products has been reported for both species [25,26]. This richness in secondary metabolites and partially overlapping chemical profiles (e.g., atranorin and chloroatranorin) make these species particularly interesting study systems for combining genome mining and comparative genomics to investigate biosynthetic enzymes that may be responsible for the production of characteristic lichen substance classes. Moreover, both species are harvested in large quantities for the perfume industry and are of economic importance [1,19]. A detailed list of the biosynthetic gene clusters detected in both species can be found in Tables S2 and S3.

For the gene cluster comparison, we functionally annotated 887 cluster genes of *E. prunastri* and 548 cluster genes of *P. furfuracea* with gene ontology terms and protein names. The full Blast2GO annotation is presented in Tables S4 and S5. We then identified 126 orthologous pairs as Reciprocal Best Blast Hits (RBH) between the *E. prunastri* and *P. furfuracea* cluster genes. The full list of RBHs is given in Table S6. From these results we chose all biosynthetic gene clusters that contained an orthologous core PKS, NRPS or hybrid PKS-NRPS for our comparative genetic mapping approach presented in the following synteny plots of Figures 3–5. The phylogenetic placements of these clusters containing a ketosynthase (KS) domain are presented below.

Four NR-PKS clusters show a core RBH gene (Figure 3). Three of these (Figure 3A,B,D) show a high cluster homology between *E. prunastri* and *P. furfuracea* with several genes carrying similar functional annotation and RBHs of each other. The homologous genes include for example cytochrome P450, monooxygenases, serine/threonine kinases, and regulatory genes. Most of the homologous regions between the clusters are confined within genes but there are a few exceptions (Figure 3D). These exceptions with high homology of coding sequences to non-coding regions may be an artefact of gene prediction, and may indicate genes missed in the annotation. The clusters in Figure 3C on the other hand display only similarity between the two core PKS genes—one of which is a R-PKSs—and not between any accessory genes. Closer inspection of the domain architecture of the NR-PKSs in Figure 3 using BLASTp indicates the presence of acyl carrier protein (ACP) transacylase starter units (SAT) and product templates (PT). Both conserved domains were shown to be typical for NR-PKSs in addition to the minimal PKS domain structure of a KS, AT (acyltransferase) and ACP [65–68].

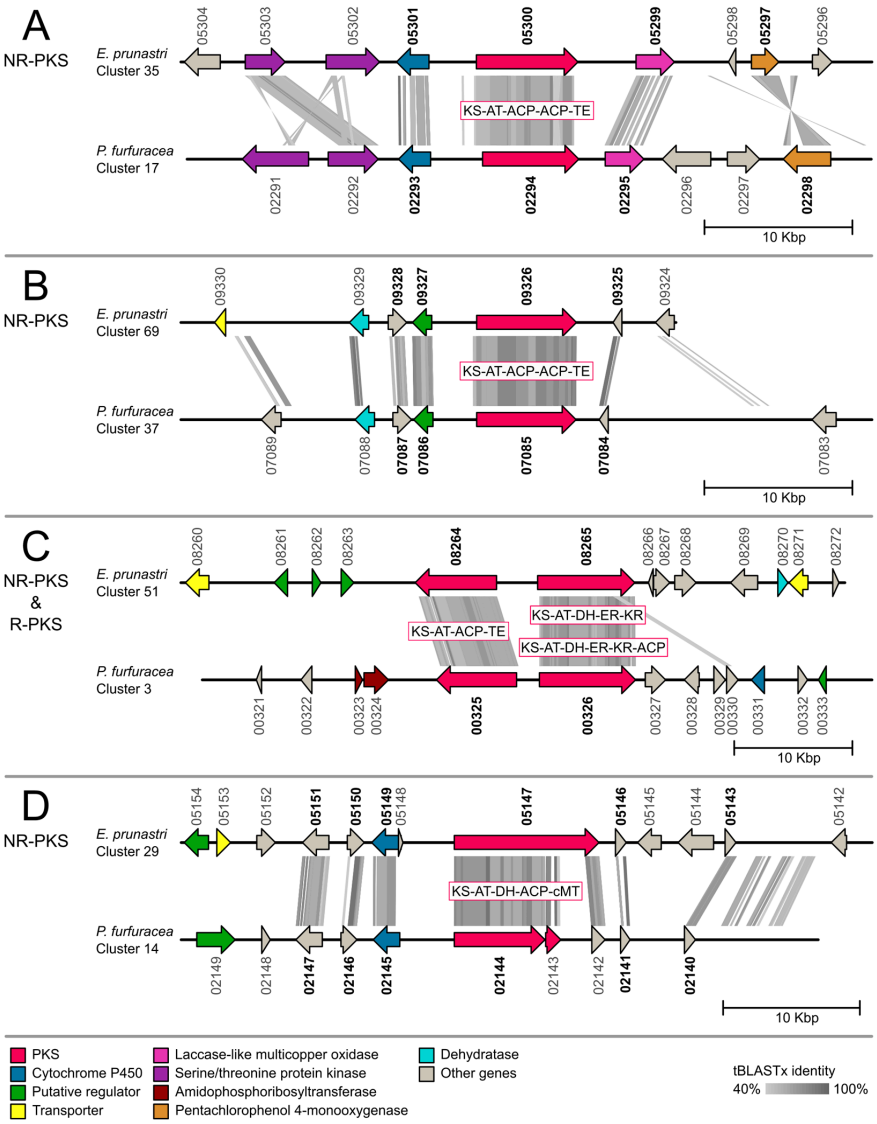


Figure 3. Synteny plots of biosynthetic gene clusters with an ortholog non-reducing polyketide synthase (NR-PKS) core gene in *Evernia prunastri* and *Pseudevernia furfuracea*. Orthologous genes identified with the reciprocal best hit (RBH) approach are highlighted in bold. ACP = Acyl carrier protein; AT = Acyltransferase; cMT = C-Methyltransferase; DH = Dehydratase; ER = Enoylreductase; KR = Ketoreductase; KS = Ketosynthase; TE = Thioesterase.

We further identified four R-PKS clusters with orthologous core genes (Figure 4). Two cluster comparisons (Figure 4A,B) show a high homology between both species. The entire R-PKS in Figure 4A is split over two to three genes—possibly an artefact of gene prediction—but we detected nine RBH accessory genes in the clusters including oxidases and calcium-binding domain proteins. The homologous cluster genes in Figure 4B include cytochrome P450, dehydrogenases, transporter genes and chalcone synthases (type III PKS). The other two comparative mapped clusters (Figure 4C,D)

share only the core gene as a RBH. In Figure 4C we detected a homologous region with annotated genes in *P. furfuracea* while *E. prunastri* lacked any predicted gene models in the concurrent region.

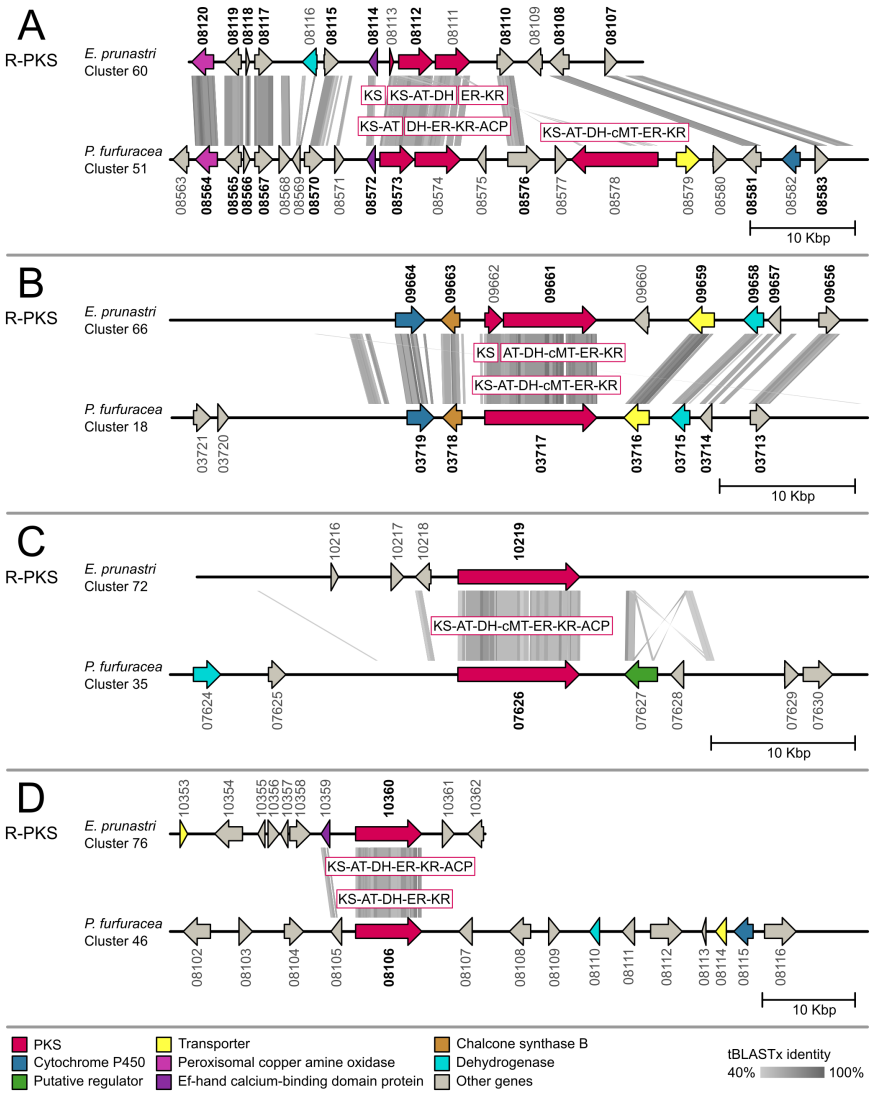


Figure 4. Synteny plots of biosynthetic gene clusters with an ortholog reducing polyketide synthase (R-PKS) core gene in *Evernia prunastri* and *Pseudevernia furfuracea*. Orthologous genes identified with the reciprocal best hit (RBH) approach are highlighted in bold. ACP = Acyl carrier protein; AT = Acyltransferase; cMT = C-Methyltransferase; DH = Dehydratase; ER = Enoylreductase; KR = Ketoreductase; KS = Ketosynthase; TE = Thioesterase.

Last, we included NRPS and hybrid PKS-NRPS clusters with orthologous core genes in the comparative analyses (Figure 5). We detected two hybrid PKS-NRPS clusters (Figure 5A,C) and two NRPS clusters (Figure 5B,C) all showing a high similarity in functionally annotated genes and through RBHs. Accessory genes include transporter genes, hydrolases and putative regulatory genes. In Figure 5C a hybrid PKS-NRPS and a NRPS are encoded closely together. This might suggest that

both enzymes participate in the biosynthesis of one natural product or that the two natural products might function together.

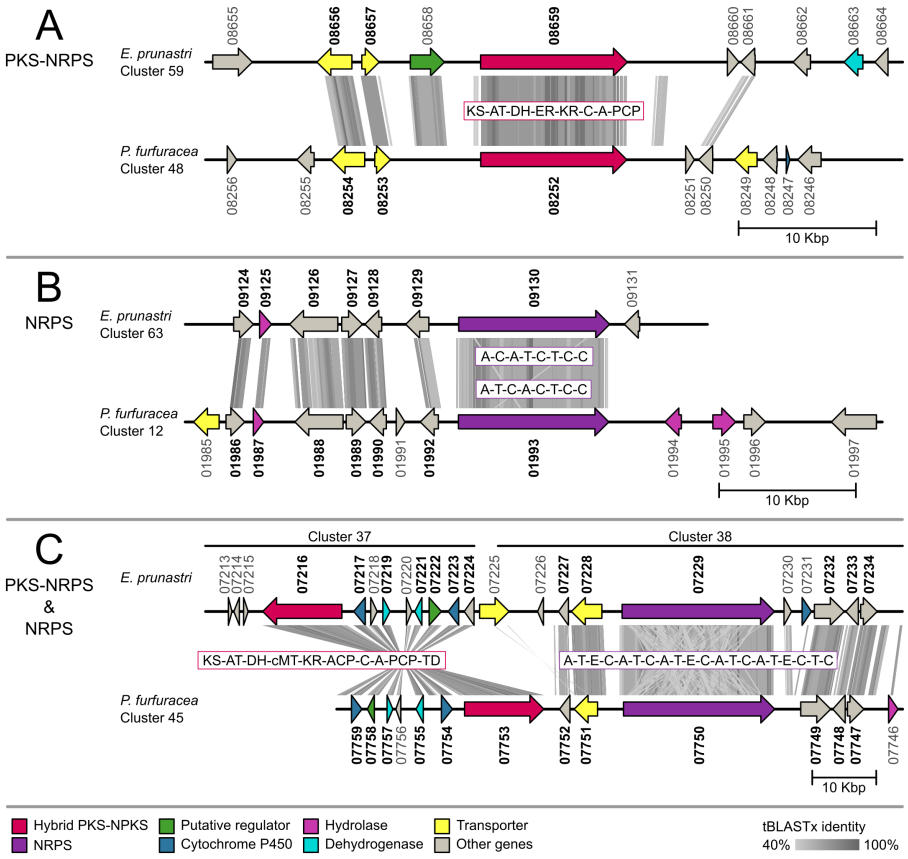


Figure 5. Synteny plots of biosynthetic gene clusters with an ortholog non-ribosomal peptide synthetase (NRPS) core gene or hybrid with polyketide synthase (PKS-NRPS) in *Evernia prunastri* and *Pseudevernia furfuracea*. Orthologous genes identified with the reciprocal best hit (RBH) approach are highlighted in bold. A = Adenylation; ACP = Acyl carrier protein; AT = Acyltransferase; C = Condensation; cMT = C-Methyltransferase; DH = Dehydratase; E = Epimerization; ER = Enoylreductase; KR = Ketoreductase; KS = Ketosynthase; PCP = peptide carrier protein; T = Thiolation; TD = Terminal domain.

2.3. Putative Regulators of Biosynthetic Gene Clusters

We identified a total of 60 putative intracluster regulators in the biosynthetic gene clusters of *E. prunastri* and *P. furfuracea*. Most of these were present in the gene clusters of *E. prunastri* (44 genes). Two pairs are orthologs between both species based on RBH analysis and can be found among the clusters investigated with comparative mapping (see clusters in Figures 3B and 5C). We detected ten C6 zinc finger domain proteins, six Zn_2Cys_6 DNA-binding proteins and two ankyrin repeat proteins among other fungal specific transcription factors. These are typical regulators of secondary metabolites [37,69]. A detailed list of these putative regulatory genes can be found in Table S7.

2.4. Phylogenetic Analysis with Characterized Fungal Polyketides

We reconstructed a maximum likelihood tree of the conserved KS domain that allows inference of domain architecture and pathway association of PKSs [52,70,71]. We included a total of 413

KS sequences from fifteen genomes of lichen-forming fungi in relation to 131 MIBiG entries of characterized fungal PKSs and six animal fatty acid synthases (FAS) as an outgroup. We also included partial PKSs from lichen-forming fungi that lack one of the required domains AT or ACP for a minimal PKS organization. Out of 413 lichen KS sequences five originate from a PKS without an AT domain and 136 from a PKS without an ACP domain. We included these since such partial PKSs are also present in the MIBiG repository of characterized enzymes [72] and may nevertheless represent functional and interesting genes. In our manually curated MIBiG dataset we found one entry without an AT domain and 32 entries without an ACP domain.

The entire phylogenetic tree is shown in full length in Supplementary Figure S1. Overall the phylogeny shows support for a clade with NR-PKSs, a clade with R-PKS and a clade containing 6-methylsalicyclic acid synthases (6-MSAS). Hybrid PKS-NRPS genes contain reducing domains and fall within the clade of R-PKS. We provide here a few examples of interesting clades found in the phylogenetic analysis with a focus on NR-PKS since the basic building subunits for lichen-specific depsides and depsidones are believed to be encoded by this gene family [1,14,27,73].

Within the NR-PKS clade we find one supported clade of PKSs containing C-methyltransferase (cMT) domains including a subclade with the PKS proposed to produce the lichen substance usnic acid [41]. *Evernia prunastri* is also known to produce usnic acid (Table 1) and one PKS (gene 02873) falls phylogenetically close to the potential usnic acid producer (Figure 6).

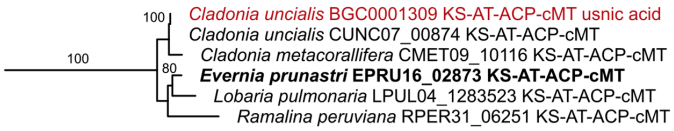


Figure 6. Supported clade containing the gene *MPAS* (methylphloroacetophenone synthase) of *Cladonia uncialis* (MIBiG-ID BGC0001309) putatively associated with usnic acid biosynthesis. For the complete KS tree see Figure S1.

Another clade contains the NR-PKSs of *E. prunastri* and *P. furfuracea* (Figure 3A) which are phylogenetically close to the PKS gene that is putatively associated with grayanic acid biosynthesis [43]. All *Cladonia* species included in this study (*C. grayi*, *C. macilenta*, *C. metacorallifera* and *C. uncialis*) have members within this clade (Figure 7).

Figure 7. Supported clade containing the gene *PKS16* of *Cladonia grayi* (MIBiG-ID BGC0001266), putatively involved in grayanic acid biosynthesis. The cluster 35 of *Evernia prunastri* (EPRU35) and cluster 17 of *Pseudevernia prunastri* (PFUR17) are phylogenetically close and are presented in detail in synteny plot Figure 3A. For the complete KS tree see Figure S1.

We further inspected the phylogenetic placement of the comparative mapped NR-PKSs of the *E. prunastri* and *P. furfuracea* gene clusters. The gene clusters mapped in Figure 3B fall into a group with tandem ACP domains and in the proximity of characterized PKSs involved in the production of hydroxy naphthalenes [74,75] (Figure 8). Naphthalene is found in the extracts of *E. prunastri* and *P. furfuracea* used for the perfume industry [1] (pp. 132–133).

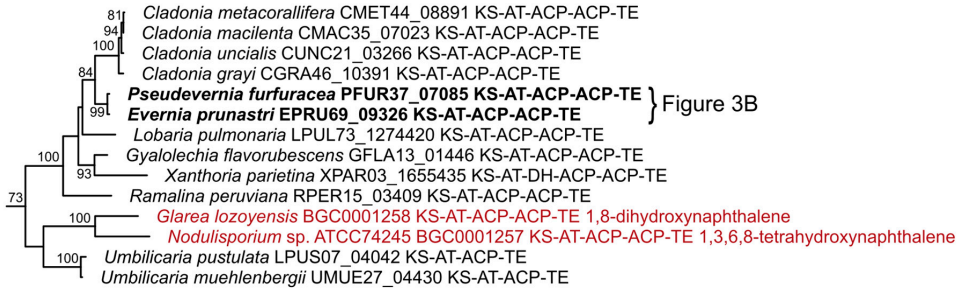


Figure 8. Supported clade containing PKSs from dihydroxy and tetrahydroxy naphthalene producers (MIBiG-ID BGC0001257, BGC0001258). For the complete KS tree see Figure S1.

Two closely related clusters of *E. prunastri* and *P. furfuracea* (Figure 3C) fall into a clade with the experimentally characterized PKS gene *orsA* known to be involved in orsellinic acid biosynthesis in *Aspergillus nidulans* [76]. This clade (Figure 9) provides an especially interesting starting point for further investigations as the most abundant class of lichen metabolites are composed of orcinol or β-orcinol monomers [27].

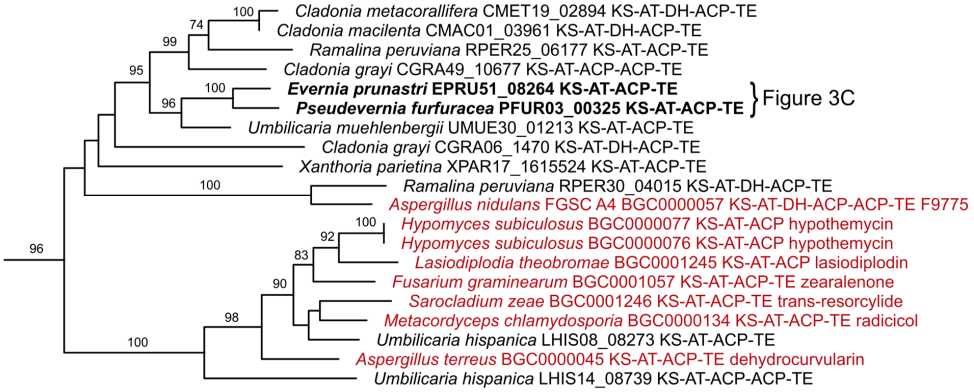


Figure 9. Supported clade containing the PKS *orsA* for orsellinic acid in *Aspergillus nidulans* (MIBiG-ID BGC0000057). For the complete KS tree see Figure S1.

We also want to underline that combining comparative mapping of entire gene clusters with a phylogenetic approach based only on the KS sequence of the core PKS gene shows that entire cluster similarities correspond to KS topology. This is in line with findings that emphasize the predictive power of the conserved KS domains for the investigation of enzyme architecture and pathway association [52,70,71].

Our study combines genome mining and comparative genomics and highlights the high diversity of biosynthetic gene clusters that can be found in fifteen genomes of lichen-forming fungi. This number exceeds by far the number of lichen metabolites that are reported for these species. The secondary metabolite gene cluster richness found in the genomes of lichen-forming fungi is above the average richness found in other representative fungal species, especially polyketide synthase gene clusters. The comparative mapping of interesting biosynthetic gene clusters, functional annotation of accessory genes together with the identification of putative regulatory genes presented here will aid in providing new insights into lichen biosynthetic pathways and serve as a valuable resource for developing heterologous expression of lichen biosynthetic gene clusters.

3. Materials and Methods

3.1. Identification and Annotations of Biosynthetic Gene Clusters

We used the genomes of the two lichen-forming fungi of *Evernia prunastri* and *Pseudevernia furfuracea* for biosynthetic gene cluster identification (accession numbers in Table 1). Sequencing of cultures, genome assemblies, gene prediction and genome mining for biosynthetic gene clusters in both species were done as described in Meiser et al. [46]. We provide here a short description on the identification of biosynthetic gene clusters to give a full depiction on our focused genome mining approach. We identified typical gene families of secondary metabolism and annotated biosynthetic gene clusters with the fungal version of antiSMASH v.4.0.2 (fungiSMASH) [62,77], including polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS) and terpene synthases. These gene families are typical targets in lichen-metabolite genome mining, because they encode the structural frame of most secondary metabolites [19,37,41,44]. As input for antiSMASH we used an annotated nucleotide file (EMBL format) constructed from the genome FASTA file and the GFF file from gene prediction.

Both investigated species show a high number of biosynthetic gene clusters with reducing and non-reducing PKSs, NRPSs, and terpene synthases [46] and represent a promising source of bioactive secondary metabolites. To further investigate these biosynthetic gene clusters for accessory genes involved in the lichen metabolite synthesis, we annotated all predicted protein sequences belonging to an antiSMASH cluster with gene ontology (GO) terms and protein names. The functional annotation was done using Blast2GO [78] v.5.0.22 and by running BLASTp [79] v.2.2.29 against the NCBI GenBank protein database *nr* (downloaded May 2017) [80] and InterProScan [81] v.5.28–67.0 with a matching lookup service and the Pfam (protein family) database v.31.0 [82].

3.2. Gene Cluster Comparison in Evernia prunastri and Pseudevernia furfuracea

We applied a two-step approach to compare the biosynthetic genes cluster of *E. prunastri* and *P. furfuracea*. First, as a simple and fast way of detecting orthologous pairs [83], we identified Reciprocal Best Blast Hits (RBH) between all cluster genes of both species. For better detecting orthologs we used BLASTp v.2.2.31 + with final Smith-Waterman alignment and soft filtering (BLAST flags *use_sw_tback*, *soft_masking true*, *seq yes* and *evalue 1e-6*) [83,84]. The BLAST hits were then filtered for a minimum query coverage of 50% (*qcovhsp*) and a minimum identity of 70% over the alignment length (*pident*) and sorted for highest bit-score and lowest e-value [85].

Second, we used comparative genetic mapping to analyse the homology and arrangement of entire clusters instead of looking at core gene orthologs only. For the comparison we chose all biosynthetic gene clusters that contained a RBH core PKS, NRPS, or hybrid PKS-NRPS. Synteny plots were generated with Easyfig python script v.2.2.3 [86] and the tBLASTx option with a minimum identity value of 40% and a minimum alignment length of 50 [87]. Gene clusters were reversed when necessary to have a matching orientation. We generated GBK files with seqkit [88] v.0.7.2 and the seqret tool in the EMBOSS package [89] v6.6.0.0 as input for Easyfig.

3.3. Putative Regulators of Biosynthetic Gene Clusters

The transcriptional control of biosynthetic genes by transcription factors is complex and can occur on several levels from pathway-specific to a broader global regulation [37]. Specific acting regulatory genes are usually found in the cluster that the factors regulate. Examples include $Zn(II)_2Cys_6$, Cys_2His_2, basic leucine zipper (bZip), winged helix, zinc-finger, or ankyrin repeat proteins [30,69]. The identification of these regulators involved in the synthesis of secondary metabolites may be crucial for successful heterologous expression experiments or for activating silent gene clusters.

For the detection of putative regulators within the biosynthetic gene clusters of *E. prunastri* and *P. furfuracea* we made use of the Blast2GO annotation. We searched for the following key terms in the GO names: 'signal transducer activity', 'transcription factor activity' and 'transcription factor binding'.

3.4. Biosynthetic Gene Richness in Other Genomes of Lichenized and Non-Lichenized Fungi

We assessed the chemical diversity and biosynthetic gene richness of *E. prunastri* and *P. furfuracea* in comparison to the genomes of other lichen-forming fungi. We analysed all genomes of lichen-forming fungi from fungal culture available in NCBI (National Center for Biotechnology Information) and for authorized genomes in JGI (DOE Joint Genome Institute) (Table 1). The only genome not derived from fungal culture is the genome of *Umbilicaria pustulata* where the fungal genome was reconstructed with metagenomic sequencing of the lichen thallus instead [90]. We ran antiSMASH as described above on all thirteen additional genomes.

For seven of these genomes no gene set was available and gene models had to be predicted before running antiSMASH to improve the biosynthetic gene cluster detection. The de novo gene prediction and annotation of these genomes was done with MAKER [91] v2.31.8 using an iterative approach as recommended in the protocols of Campbell et al. [92]. In the first round of MAKER we used Hidden Markov Models (HMMs) generated with GeneMark-ES [93] v4.33 and SNAP [94] with hints from CEGMA [95] v2.4 (performed on iPlant [96]) as training data for gene finding. The first-round results were then converted to new SNAP and Augustus [97] v3.0.2 HMMs for the second round of MAKER resulting in the final set of gene models.

The genome and gene set completeness of all 15 lichen genomes was assessed based on evolutionarily-informed expectations of gene content with BUSCO (Benchmarking Universal Single-Copy Orthologs) [98] v.3.0 and a lineage-specific set of 1315 Ascomycota single-copy orthologs from OrthoDB [99] v.9. Completeness estimates were visualized with the package ggplot2 [100] v.2.2.1 in R [101] v.3.4.4.

Additionally, we provide an overview on how richness of biosynthetic gene clusters found in lichenized fungi compare to the cluster richness found in non-lichenized fungi. Therefore, we combined the secondary metabolic gene cluster counts for the 15 genomes of lichenized fungi obtained from antiSMASH in this study with a recently published analysis from 57 genomes of representative fungal ascomycete and basidiomycete species [32].

3.5. Phylogenetic Analysis with Characterized Fungal Polyketides

We used phylogenetics to analyse the core enzymes of PKS and hybrid PKS-NRPS gene clusters in relation to characterized fungal biosynthetic gene clusters. Combining genome mining with phylogenetic analysis may provide clues for identifying promising clusters for the production of characteristic substance classes through phylogenetic relatedness to previously characterized genes [36,62,102].

Our phylogenetic approach is based on protein sequences of the ketosynthase (KS) domain of PKSs and hybrid PKS-NRPSs. The KS domains are the most conserved and essential part of a PKS cluster and are highly predictive of enzyme architecture and pathway association [52,70,71]. We included all KS sequences identified in the fifteen genomes of lichen-forming fungi (Table 2) that stem from genes with at least three PKS domains and with KS sequences longer than 200 bases. This resulted in a dataset of 413 KS sequences originating from lichen-forming fungi. Additionally, we made use of the MIBiG repository (Minimum Information about Biosynthetic Gene cluster [103]) to include experimentally characterized biosynthetic gene clusters [72]. We compiled a custom database with 131 entries of all fungal PKS records in MIBiG (downloaded January 2018). Further we included six animal fatty acid synthase (FAS) protein sequences as outgroup for the phylogenetic inference (NCBI Reference Sequence accessions: *Bos taurus* NP_001012687.1; *Gallus gallus* NP_990486.2; *Homo sapiens* NP_004095.4; *Mus musculus* NP_032014.3; *Rattus norvegicus* NP_059028.1; *Sus scrofa* NP_001093400.1).

We ran MAFFT [104,105] v7.309 in Geneious v9.1.8 (https://www.geneious.com) to generate a multiple sequence alignment of all 550 KS amino acid sequences resulting in an alignment with a length of 1043 characters. We chose the iterative refinement algorithm G-INS-i with a BLOSUM62 scoring matrix, a gap open penalty of 1.53 and an offset value of 0.123. We performed ProtTest [106] v.3.4.2 on our alignment to select the best-fitting substitution model of amino acid replacement for

subsequent tree reconstruction. Model-testing indicated LG + I + G as the best model for both the Akaike and the Bayesian Information Criterion. The phylogenetic tree was then generated with RAxML-HPC Black Box v.8.2.10 (Randomized Axelerated Maximum Likelihood) [107] on CIPRES v3.3 (Cyperinfrastructure for Phylogenetic Research) [108,109] with automated bootstrapping, protein sequence type, FAS sequences set as outgroup, estimating proportion of invariable sites (GTRGAMMA + I), protein substitution matrix LG, no empirical base frequencies and printing branch lengths. The tree was visualized with Figtree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree) rooting the tree with the help of the FAS outgroup and ordering nodes increasingly.

Supplementary Materials: The following are available online, Table S1: Genome Statistics and Completeness of Lichen-Forming Fungi; Table S2: Overview of Biosynthetic Gene Clusters in *Evernia prunastri*; Table S3: Overview of Biosynthetic Gene Clusters in *Pseudevernia prunastri*; Table S4: Functional Annotation of *Evernia prunastri* Cluster Genes; Table S5: Functional Annotation of *Pseudevernia furfuracea* Cluster Genes; Table S6: Reciprocal Best Blast Hits of *Evernia prunastri* and *Pseudevernia furfuracea* Cluster Genes; Table S7: List of Putative Regulatory Genes in *Evernia prunastri* and *Pseudevernia furfuracea*; Figure S1: Phylogenetic Tree Based on the Ketosynthase (KS) Domain; Supplementary File S1: Ketosynthase (KS) Alignment for Phylogenetic Tree Reconstruction (FASTA); Supplementary Files S2: Predicted Gene Models for Seven Genomes of Lichen-Forming Fungi (GFF3); All output of antiSMASH analyses are available upon request from the corresponding author.

Author Contributions: Conceptualization, A.C., F.D.G. and I.S.; formal analysis, A.C.; investigation A.C.; methodology A.C.; validation, F.D.G., I.S. and H.B.B.; writing—original draft preparation, A.C.; visualization, A.C.; writing—review and editing, F.D.G., I.S. and H.B.B.; supervision, F.D.G. and I.S.; project administration, I.S.

Funding: This research was supported by the research funding program Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz (LOEWE) of Hesse's Ministry of Higher Education, Research and the Arts through the Priority Programme Integrative Pilzforschung (IPF) and the Translational Biodiversity Genomics (TBG) Center. Additional funding came from the RMU Initiativfond "Biotic interactions". Some analyses were performed on the FUCHS cluster of the Center for Scientific Computing (CSC) of the Goethe University in Frankfurt am Main.

Acknowledgments: We thank the following colleagues for permission to use genomes prior to full publication: Daniele Armaleo (Duke University)—*Cladonia grayi*, Ólafur S. Andrésson (University of Iceland)—*Lobaria pulmonaria* and Paul Dyer (University of Nottingham)—*Xanthoria parietina*.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nash, T.H. *Lichen Biology*, 2nd ed.; Nash, T.H., Ed.; Cambridge University Press: Cambridge, UK, 2008; ISBN 9780521871624.
2. Lücking, R.; Hodkinson, B.P.; Leavitt, S.D. The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota—Approaching one thousand genera. *Bryologist* **2016**, *119*, 361–416. [CrossRef]
3. Asplund, J.; Wardle, D.A. How lichens impact on terrestrial community and ecosystem properties. *Biol. Rev.* **2017**, *92*, 1720–1738. [CrossRef] [PubMed]
4. Seaward, M.R.D. Environmental role of lichens. In *Lichen Biology*; Nash, T.H., Ed.; Cambridge University Press: Cambridge, UK, 2008; pp. 274–298. ISBN 9780521871624.
5. Grube, M.; Wedin, M. Lichenized fungi and the evolution of symbiotic organization. *Microbiol. Spectr.* **2016**, *4*, 1–17. [CrossRef]
6. Cernava, T.; Berg, G.; Grube, M. High life expectancy of bacteria on lichens. *Microb. Ecol.* **2016**, *72*, 510–513. [CrossRef] [PubMed]
7. Aschenbrenner, I.A.; Cernava, T.; Berg, G.; Grube, M. Understanding microbial multi-species symbioses. *Front. Microbiol.* **2016**, *7*, 180. [CrossRef] [PubMed]
8. Spribille, T.; Tuovinen, V.; Resl, P.; Vanderpool, D.; Wolinski, H.; Aime, M.C.; Schneider, K.; Stabentheiner, E.; Toome-Heller, M.; Thor, G.; et al. Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **2016**, *353*, 488–492. [CrossRef] [PubMed]
9. Fernández-Mendoza, F.; Fleischhacker, A.; Kopun, T.; Grube, M.; Muggia, L. ITS1 metabarcoding highlights low specificity of lichen mycobiomes at a local scale. *Mol. Ecol.* **2017**, *26*, 4811–4830. [CrossRef] [PubMed]

10. Dal Grande, F.; Rolshausen, G.; Divakar, P.K.; Crespo, A.; Otte, J.; Schleuning, M.; Schmitt, I. Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol.* **2018**, *217*, 277–289. [CrossRef]
11. Machado, H.; Tuttle, R.N.; Jensen, P.R. Omics-based natural product discovery and the lexicon of genome mining. *Curr. Opin. Microbiol.* **2017**, *39*, 136–142. [CrossRef] [PubMed]
12. Cernava, T.; Erlacher, A.; Aschenbrenner, I.A.; Krug, L.; Lassek, C.; Riedel, K.; Grube, M.; Berg, G. Deciphering functional diversification within the lichen microbiota by meta-omics. *Microbiome* **2017**, *5*, 82. [CrossRef]
13. Moya, P.; Molins, A.; Martínez-Alberola, F.; Muggia, L.; Barreno, E. Unexpected associated microalgal diversity in the lichen *Ramalina farinacea* is uncovered by pyrosequencing analyses. *PLoS ONE* **2017**, *12*, e0175091. [CrossRef] [PubMed]
14. Stocker-Wörgötter, E. Metabolic diversity of lichen-forming ascomycetous fungi: Culturing, polyketide and shikimatemetabolite production, and PKS genes. *Nat. Prod. Rep. Prod. Rep.* **2008**, *25*, 188–200. [CrossRef] [PubMed]
15. Molnár, K.; Farkas, E. Current results on biological activities of lichen secondary metabolites: A review. *Z. Naturforsch. C* **2010**, *65*, 157–173. [CrossRef] [PubMed]
16. Elix, J.A.; Stocker-Wörgötter, E. Biochemistry and secondary metabolites. In *Lichen Biology*; Nash, T.H., Ed.; Cambridge University Press: Cambridge, UK, 2008; pp. 104–133. ISBN 9780521871624.
17. Huneck, S.; Yoshimura, I. *Identification of Lichen Substances*; Springer: Berlin/Heidelberg, Germany, 1996; ISBN 978-3-642-85245-9.
18. Rundel, P.W. The ecological role of secondary lichen substances. *Biochem. Syst. Ecol.* **1978**, *6*, 157–170. [CrossRef]
19. Huneck, S. The significance of lichens and their metabolites. *Naturwissenschaften* **1999**, *86*, 559–570. [CrossRef] [PubMed]
20. Lawrey, J.D. Biological role of lichen substances. *Bryologist* **1986**, *89*, 111. [CrossRef]
21. Ranković, B.; Kosanić, M. *Lichen Secondary Metabolites*; Ranković, B., Ed.; Springer: Cham, Switzerland, 2015; ISBN 978-3-319-13373-7.
22. Boustie, J.; Grube, M. Lichens—A promising source of bioactive secondary metabolites. *Plant Genet. Resour. Charact. Util.* **2005**, *3*, 273–287. [CrossRef]
23. Zhou, R.; Yang, Y.; Park, S.-Y.; Nguyen, T.T.; Seo, Y.-W.; Lee, K.H.; Lee, J.H.; Kim, K.K.; Hur, J.-S.; Kim, H. The lichen secondary metabolite atranorin suppresses lung cancer cell motility and tumorigenesis. *Sci. Rep.* **2017**, *7*, 8136. [CrossRef]
24. Crawford, S.D. Lichens used in traditional medicine. In *Lichen Secondary Metabolites*; Springer: Cham, Switzerland, 2015; pp. 27–80. ISBN 9783319133744.
25. Joulain, D.; Tabacchi, R. Lichen extracts as raw materials in perfumery. Part 1: Oakmoss. *Flavour Fragr. J.* **2009**, *24*, 49–61. [CrossRef]
26. Joulain, D.; Tabacchi, R. Lichen extracts as raw materials in perfumery. Part 2: Treemoss. *Flavour Fragr. J.* **2009**, *24*, 105–116. [CrossRef]
27. Calcott, M.J.; Ackerley, D.F.; Knight, A.; Keyzers, R.A.; Owen, J.G. Secondary metabolism in the lichen symbiosis. *Chem. Soc. Rev.* **2018**, *47*, 1730–1760. [CrossRef] [PubMed]
28. Crawford, J.M.; Townsend, C.A. New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* **2010**, *8*, 879–889. [CrossRef] [PubMed]
29. Schmitt, I.; Martín, M.P.; Kautz, S.; Lumbsch, H.T. Diversity of non-reducing polyketide synthase genes in the Pertusariales (lichenized Ascomycota): A phylogenetic perspective. *Phytochemistry* **2005**, *66*, 1241–1253. [CrossRef] [PubMed]
30. Keller, N.P.; Turner, G.; Bennett, J.W. Fungal secondary metabolism—From biochemistry to genomics. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947. [CrossRef] [PubMed]
31. Cox, R.J.; Skellam, E.; Williams, K. Biosynthesis of fungal polyketides. In *Physiology and Genetics*; Springer: Cham, Switzerland, 2018; pp. 385–412.
32. Rokas, A.; Wisecaver, J.H.; Lind, A.L. The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* **2018**, *16*, 731–744. [CrossRef] [PubMed]
33. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Wieland Brown, L.C.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421. [CrossRef]

34. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.-S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968. [CrossRef]
35. Zhao, H.; Medema, M.H. Standardization for natural product synthetic biology. *Nat. Prod. Rep.* **2016**, *33*, 920–924. [CrossRef]
36. Nielsen, J.C.; Grijseels, S.; Prigent, S.; Ji, B.; Dainat, J.; Nielsen, K.F.; Frisvad, J.C.; Workman, M.; Nielsen, J. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat. Microbiol.* **2017**, *2*, 17044. [CrossRef]
37. Brakhage, A.A. Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **2013**, *11*, 21–32. [CrossRef]
38. Schümann, J.; Hertweck, C. Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J. Biotechnol.* **2006**, *124*, 690–703. [CrossRef]
39. Hertweck, C. Hidden biosynthetic treasures brought to light. *Nat. Chem. Biol.* **2009**, *5*, 450–452. [CrossRef] [PubMed]
40. Bergmann, S.; Schümann, J.; Scherlach, K.; Lange, C.; Brakhage, A.A.; Hertweck, C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat. Chem. Biol.* **2007**, *3*, 213–217. [CrossRef] [PubMed]
41. Abdel-Hameed, M.; Bertrand, R.L.; Piercey-Normore, M.D.; Sorensen, J.L. Putative identification of the usnic acid biosynthetic gene cluster by de novo whole-genome sequencing of a lichen-forming fungus. *Fungal Biol.* **2016**, *120*, 306–316. [CrossRef] [PubMed]
42. Muggia, L.; Schmitt, I.; Grube, M. Lichens as treasure chests of natural products. *Sim News* **2009**, *59*, 85–97.
43. Armaleo, D.; Sun, X.; Culberson, C. Insights from the first putative biosynthetic gene cluster for a lichen depside and depsidone. *Mycologia* **2011**, *103*, 741–754. [CrossRef]
44. Abdel-Hameed, M.; Bertrand, R.L.; Piercey-Normore, M.D.; Sorensen, J.L. Identification of 6-hydroxymellein synthase and accessory genes in the lichen *Cladonia uncialis*. *J. Nat. Prod.* **2016**, *79*, 1645–1650. [CrossRef]
45. Bertrand, R.L.; Sorensen, J.L. A comprehensive catalogue of polyketide synthase gene clusters in lichenizing fungi. *J. Ind. Microbiol. Biotechnol.* **2018**, *45*, 1067–1081. [CrossRef]
46. Meiser, A.; Otte, J.; Schmitt, I.; Dal Grande, F. Sequencing genomes from mixed DNA samples—Evaluating the metagenome skimming approach in lichenized fungi. *Sci. Rep.* **2017**, *7*, 14881. [CrossRef]
47. Wang, Y.; Geng, C.; Yuan, X.; Hua, M.; Tian, F.; Li, C. Identification of a putative polyketide synthase gene involved in usnic acid biosynthesis in the lichen *Nephromopsis pallescens*. *PLoS ONE* **2018**, *13*, e0199110. [CrossRef]
48. Taguchi, H.; Sankawa, U.; Shibata, S. Biosynthesis of natural products. VI. Biosynthesis of usnic acid in lichens. (1). A general scheme of biosynthesis of usnic acid. *Chem. Pharm. Bull. (Tokyo)* **1969**, *17*, 2054–2060. [CrossRef] [PubMed]
49. Barton, D.H.R.; Deflorin, A.M.; Edwards, O.E. 108. The synthesis of usnic acid. *J. Chem. Soc.* **1956**, *108*, 530–534. [CrossRef]
50. Gagunashvili, A.N.; Davíðsson, S.P.; Jónsson, Z.O.; Andrésson, Ó.S. Cloning and heterologous transcription of a polyketide synthase gene from the lichen *Solorina crocea*. *Mycol. Res.* **2009**, *113*, 354–363. [CrossRef] [PubMed]
51. Chooi, Y.-H.; Stalker, D.M.; Davis, M.A.; Fujii, I.; Elix, J.A.; Louwhoff, S.H.J.J.; Lawrie, A.C. Cloning and sequence characterization of a non-reducing polyketide synthase gene from the lichen *Xanthoparmelia semiviridis*. *Mycol. Res.* **2008**, *112*, 147–161. [CrossRef] [PubMed]
52. Ziemert, N.; Jensen, P.R. Phylogenetic approaches to natural product structure prediction. In *Methods in Enzymology*; Elsevier Inc.: San Diego, CA, USA, 2012; Volume 517, pp. 161–182. ISBN 9780124046344.
53. Wang, Y.; Wang, J.; Cheong, Y.H.; Hur, J.-S. Three new non-reducing polyketide synthase genes from the lichen-forming fungus *Usnea longissima*. *Mycobiology* **2014**, *42*, 34–40. [CrossRef] [PubMed]
54. Bertrand, R.L.; Abdel-Hameed, M.; Sorensen, J.L. Lichen biosynthetic gene clusters part II: Homology mapping suggests a functional diversity. *J. Nat. Prod.* **2018**, *81*, 732–748. [CrossRef] [PubMed]
55. Purvis, O.W.; Coppins, B.J.; Hawksworth, D.L.; James, P.W.; Moore, D.M. (Eds.) *The Lichen Flora of Great Britain and Ireland*; Natural History Museum Publications in association with The British Lichen Society: London, UK, 1992; ISBN 0952304902.
56. Brodo, I.M.; Sharnoff, S.D.; Sharnoff, S. (Eds.) *Lichens of North America*; Yale University Press: New Haven, CO, USA; London, UK, 2001; ISBN 0-300-08249-5.

57. Stenroos, S.; Velmala, S.; Pykälä, J.; Ahti, T. (Eds.) *Lichens of Finland*; Finnish Museum of Natural History LUMOS, University of Helsinki: Helsinki, Finland, 2016; ISBN 978-951-51-2266-7.
58. Posner, B.; Feige, G.B.; Leuckert, C. Beiträge zur Chemie der Flechtengattung *Lasallia* Mérat/On the Chemistry of the lichen genus *Lasallia* Merat. *Z. Naturforsch. C* **1991**, *46*, 19–27. [CrossRef]
59. Dal Grande, F.; Meiser, A.; Greshake Tzovaras, B.; Otte, J.; Ebersberger, I.; Schmitt, I. The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *Lichenologist* **2018**, *50*, 329–340. [CrossRef]
60. Park, S.-Y.; Choi, J.; Lee, G.-W.; Park, C.-H.; Kim, J.A.; Oh, S.-O.; Lee, Y.-H.; Hur, J.-S. Draft Genome Sequence of *Endocarpon pusillum* Strain KoLRILF000583. *Genome Announc.* **2014**, *2*, e00452-14. [CrossRef]
61. Wang, Y.-Y.; Liu, B.; Zhang, X.-Y.; Zhou, Q.-M.; Zhang, T.; Li, H.; Yu, Y.-F.; Zhang, X.-L.; Hao, X.-Y.; Wang, M.; et al. Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genom.* **2014**, *15*, 34. [CrossRef]
62. Blin, K.; Kim, H.U.; Medema, M.H.; Weber, T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinform.* **2017**, *bbx146*, 1–11. [CrossRef] [PubMed]
63. Stocker-Wörgötter, E. Biochemical diversity and ecology of lichen-forming fungi: Lichen substances, chemosyndromic variation and origin of polyketide-type metabolites (biosynthetic pathways). In *Recent Advances in Lichenology*; Springer: New Delhi, India, 2015; pp. 161–179. ISBN 9788132222354.
64. Wang, X.; Zhang, X.; Liu, L.; Xiang, M.; Wang, W.; Sun, X.; Che, Y.; Guo, L.; Liu, G.; Guo, L.; et al. Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. *BMC Genom.* **2015**, *16*, 28. [CrossRef]
65. Crawford, J.M.; Dancy, B.C.R.; Hill, E.A.; Udwary, D.W.; Townsend, C.A. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16728–16733. [CrossRef]
66. Crawford, J.M.; Thomas, P.M.; Scheerer, J.R.; Vagstad, A.L.; Kelleher, N.L.; Townsend, C.A. Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, *320*, 243–246. [CrossRef]
67. Crawford, J.M.; Korman, T.P.; Labonte, J.W.; Vagstad, A.L.; Hill, E.A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C.A. Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **2009**, *461*, 1139–1143. [CrossRef]
68. Cacho, R.A.; Tang, Y.; Chooi, Y.-H. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Front. Microbiol.* **2015**, *5*, 1–16. [CrossRef] [PubMed]
69. Knox, B.P.; Keller, N.P. Key players in the regulation of fungal secondary metabolism. In *Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites, Volume 2*; Zeilinger, S., Martín, J.-F., García-Estrada, C., Eds.; Springer: New York, NY, USA, 2015; Volume 2, pp. 13–28. ISBN 978-1-4939-2530-8.
70. Kroken, S.; Glass, N.L.; Taylor, J.W.; Yoder, O.C.; Turgeon, B.G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15670–15675. [CrossRef] [PubMed]
71. Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **2005**, *22*, 2027–2039. [CrossRef]
72. Li, Y.F.; Tsai, K.J.S.; Harvey, C.J.B.; Li, J.J.; Ary, B.E.; Berlew, E.E.; Boehman, B.L.; Findley, D.M.; Friant, A.G.; Gardner, C.A.; et al. Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. *Fungal Genet. Biol.* **2016**, *89*, 18–28. [CrossRef]
73. Nguyen, K.-H.; Chollet-Krugler, M.; Gouault, N.; Tomasi, S. UV-protectant metabolites from lichens and their symbiotic partners. *Nat. Prod. Rep.* **2013**, *30*, 1490. [CrossRef]
74. Fulton, T.R.; Ibrahim, N.; Losada, M.C.; Grzegorski, D.; Tkacz, J.S. A melanin polyketide synthase (PKS) gene from *Nodulisporium* sp. that shows homology to the *pks1* gene of *Colletotrichum lagenarium*. *Mol. Gen. Genet.* **1999**, *262*, 714–720. [CrossRef] [PubMed]
75. Zhang, A.; Lu, P.; Dahl-Roshak, A.M.; Paress, P.S.; Kennedy, S.; Tkacz, J.S.; An, Z. Efficient disruption of a polyketide synthase gene (*pks1*) required for melanin synthesis through *Agrobacterium*-mediated transformation of *Glarea lozoyensis*. *Mol. Genet. Genom.* **2003**, *268*, 645–655. [CrossRef]
76. Sanchez, J.F.; Chiang, Y.-M.; Szewczyk, E.; Davidson, A.D.; Ahuja, M.; Elizabeth Oakley, C.; Woo Bok, J.; Keller, N.; Oakley, B.R.; Wang, C.C.C. Molecular genetic analysis of the orsellinic acid/F9775 genecluster of *Aspergillus nidulans*. *Mol. Biosyst.* **2010**, *6*, 587–593. [CrossRef] [PubMed]

77. Blin, K.; Wolf, T.; Chevrette, M.G.; Lu, X.; Schwalen, C.J.; Kautsar, S.A.; Suarez Duran, H.G.; de los Santos, E.L.C.; Kim, H.U.; Nave, M.; et al. antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45*, W36–W41. [CrossRef] [PubMed]
78. Conesa, A.; Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.* **2008**. [CrossRef] [PubMed]
79. Altschul, S. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
80. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2017**, *45*, D37–D42. [CrossRef]
81. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]
82. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [CrossRef]
83. Ward, N.; Moreno-Hagelsieb, G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE* **2014**, *9*, e101850. [CrossRef]
84. Moreno-Hagelsieb, G.; Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **2008**, *24*, 319–324. [CrossRef] [PubMed]
85. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef] [PubMed]
86. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [CrossRef] [PubMed]
87. Kjærbølling, I.; Vesth, T.C.; Frisvad, J.C.; Nybo, J.L.; Theobald, S.; Kuo, A.; Bowyer, P.; Matsuda, Y.; Mondo, S.; Lyhne, E.K.; et al. Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E753–E761. [CrossRef]
88. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef]
89. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [CrossRef]
90. Dal Grande, F.; Sharma, R.; Meiser, A.; Rolshausen, G.; Büdel, B.; Mishra, B.; Thines, M.; Otte, J.; Pfenninger, M.; Schmitt, I. Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evol. Biol.* **2017**, *17*, 93. [CrossRef]
91. Holt, C.; Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **2011**, *12*, 491. [CrossRef] [PubMed]
92. Campbell, M.S.; Holt, C.; Moore, B.; Yandell, M. Genome annotation and curation using MAKER and MAKER-P. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; Volume 48, pp. 4.11.1–4.11.39. ISBN 0471250953.
93. Ter-Hovhannisyan, V.; Lomsadze, A.; Chernoff, Y.O.; Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **2008**, *18*, 1979–1990. [CrossRef] [PubMed]
94. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **2004**, *5*, 59. [CrossRef] [PubMed]
95. Parra, G.; Bradnam, K.; Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **2007**, *23*, 1061–1067. [CrossRef] [PubMed]
96. Goff, S.A.; Vaughn, M.; McKay, S.; Lyons, E.; Stapleton, A.E.; Gessler, D.; Matasci, N.; Wang, L.; Hanlon, M.; Lenards, A.; et al. The iPlant collaborative: Cyberinfrastructure for plant biology. *Front. Plant Sci.* **2011**, *2*, 34. [CrossRef] [PubMed]
97. Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **2006**, *7*, 62. [CrossRef]
98. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef] [PubMed]

99. Zdobnov, E.M.; Tegenfeldt, F.; Kuznetsov, D.; Waterhouse, R.M.; Simão, F.A.; Ioannidis, P.; Seppey, M.; Loetscher, A.; Kriventseva, E.V. OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **2017**, *45*, D744–D749. [CrossRef] [PubMed]
100. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-98140-6.
101. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: https://www.R-project.org/ (accessed on 29 March 2018).
102. Bahram, M.; Hildebrand, F.; Forslund, S.K.; Anderson, J.L.; Soudzilovskaia, N.A.; Bodegom, P.M.; Bengtsson-Palme, J.; Anslan, S.; Coelho, L.P.; Harend, H.; et al. Structure and function of the global topsoil microbiome. *Nature* **2018**, *560*, 233–237. [CrossRef] [PubMed]
103. Medema, M.H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J.B.; Blin, K.; de Bruijn, I.; Chooi, Y.H.; Claesen, J.; Coates, R.C.; et al. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625–631. [CrossRef]
104. Katoh, K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
105. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
106. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [CrossRef]
107. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]
108. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*; IEEE: New Orleans, LA, USA, 2010; pp. 1–8.
109. Miller, M.A.; Schwartz, T.; Pickett, B.E.; He, S.; Klem, E.B.; Scheuermann, R.H.; Passarotti, M.; Kaufman, S.; O'Leary, M.A. A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. *Evol. Bioinform.* **2015**, *11*, EBO.S21501. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are not available from the authors.

Supplementary material to

Calchera, A., Dal Grande, F., Bode, H. B., & Schmitt, I. (2019). Biosynthetic gene content of the 'perfume lichens' *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules*, 24(1), 203, doi:10.3390/molecules24010203.

Table S1. Genome statistics and completeness of lichen-forming fungi.
Table S2. Overview of biosynthetic gene clusters in *Evernia prunastri*.[1]
Table S3. Overview of biosynthetic gene clusters in *Pseudevernia prunastri*.[1]
Table S4. Functional annotation of *Evernia prunastri* cluster genes.[2]
Table S5. Functional annotation of *Pseudevernia furfuracea* cluster genes.[2]
Table S6. Reciprocal best blast hits of *Evernia prunastri* and *Pseudevernia furfuracea*. cluster genes.
Table S7. List of putative regulatory genes in *Evernia prunastri* and *Pseudevernia furfuracea*.
Fig. S1. Phylogenetic tree based on the ketosynthase (KS) domain.
File S1. Ketosynthase (KS) alignment for phylogenetic tree reconstruction (FASTA file).[2]
File S2. Predicted gene models for seven genomes of lichen-forming Fungi (GFF3 file).[2]

[1] This table is abbreviated for this print. The full version can be found online accompanying the published paper at the address below.
[2] This supplementary material can be found online accompanying the published paper at http://dx.doi.org/10.3390/molecules24010203.

Table S1. Genome Statistics and Completeness of Lichen-Forming Fungi.

Species	Abbreviation	Genome size	Scaffold N50	Number of scaffolds	Number of contigs	BUSCO genome completeness	Number of genes	BUSCO gene set completeness	Annotation source
Cladonia grayi	CGRA	34.6 Mb	243,412	414	519	C:92.4 % [S:92.2 %, D:0.2 %], F:4.6 %, M:3.0 %, n:1,315	11,389	C:93.1 % [S:92.4 %, D:0.7 %], F:4.3 %, M:2.6 %, n:1,315	JGI
Cladonia macilenta	CMAC	37.1 Mb	1,469,036	240	1,310	C:93.1 % [S:92.8 %, D:0.3 %], F:4.3 %, M:2.6 %, n:1,315	10,559	C:80.2 % [S:80.0 %, D:0.2 %], F:13.1 %, M:6.7 %, n:1,315	this study
Cladonia metacorallifera	CMET	36.7 Mb	1,591,850	30	625	C:93.6 % [S:93.1 %, D:0.5 %], F:3.3 %, M:3.1 %, n:1,315	10,497	C:78.3 % [S:77.8 %, D:0.5 %], F:14.3 %, M:7.4 %, n:1,315	this study
Cladonia uncialis	CUNC	32.9 Mb	34,871	2,124	2,137	C:89.1 % [S:88.9 %, D:0.2 %], F:5.6 %, M:5.3 %, n:1,315	10,902	C:77.8 % [S:77.5 %, D:0.3 %], F:14.1 %, M:8.1 %, n:1,315	this study
Endocarpon pusillum (Park *et al.*)	EPUP	37.2 Mb	1,340,794	40	469	C:95.3 % [S:94.4 %, D:0.9 %], F:3.7 %, M:1.0 %, n:1,315	11,756	C:79.1 % [S:78.3 %, D:0.8 %], F:14.2 %, M:6.7 %, n:1,315	this study
Endocarpon pusillum (Wang *et al.*)	EPUW	37.1 Mb	178,225	908	1,731	C:90.9 % [S:90.9 %, D:0.0 %], F:5.2 %, M:3.9 %, n:1,315	9,238	C:89.5 % [S:89.3 %, D:0.2 %], F:4.9 %, M:5.6 %, n:1,315	NCBI
Evernia prunastri	EPRU	40.3 Mb	264,454	277	579	C:95.9 % [S:95.1 %, D:0.8 %], F:2.7 %, M:1.4 %, n:1,315	10,992	C:92.1 % [S:91.3 %, D:0.8 %], F:5.1 %, M:2.8 %, n:1,315	NCBI
Gyalolechia flavorubescens	GFLA	34.5 Mb	1,693,300	36	189	C:95.4 % [S:95.4 %, D:0.0 %], F:3.1 %, M:1.5 %, n:1,315	10,460	C:74.2 % [S:74.1 %, D:0.1 %], F:17.8 %, M:8.0 %, n:1,315	this study
Lobaria pulmonaria	LPUL	56.1 Mb	50,541	1,911	2,271	C:95.0 % [S:93.5 %, D:1.5 %], F:3.4 %, M:1.6 %, n:1,315	15,607	C:96.9 % [S:95.2 %, D:1.7 %], F:2.2 %, M:0.9 %, n:1,315	JGI
Pseudevernia furfuracea	PFUR	37.8 Mb	1,178,799	46	143	C:94.7 % [S:94.6 %, D:0.1 %], F:3.5 %, M:1.8 %, n:1,315	8,842	C:91.8 % [S:91.8 %, D:0.0 %], F:5.0 %, M:3.2 %, n:1,315	NCBI
Ramalina peruviana	RPER	27.0 Mb	40,431	1,657	37,922	C:86.3 % [S:86.1 %, D:0.2 %], F:7.0 %, M:6.7 %, n:1,315	9,338	C:72.1 % [S:71.6 %, D:0.5 %], F:15.7 %, M:12.2 %, n:1,315	this study
Umbilicaria hispanica	LHIS	41.2 Mb	145,035	1,619	3,084	C:96.4 % [S:95.6 %, D:0.8 %], F:2.1 %, M:1.5 %, n:1,315	8,488	C:82.9 % [S:82.4 %, D:0.5 %], F:2.4 %, M:14.7 %,n:1,315	NCBI
Umbilicaria muehlenbergii	UMUE	34.8 Mb	7,009,248	7	297	C:96.4 % [S:95.9 %, D:0.5 %], F:2.3 %, M:1.3 %, n:1,315	8,968	C:79.8 % [S:79.4 %, D:0.4 %], F:14.5 %, M:5.7 %, n:1,315	this study
Umbilicaria pustulata	LPUS	39.2 Mb	104,297	3,891	5,331	C:91.9 % [S:91.1 %, D:0.8 %], F:5.2 %, M:2.9 %, n:1,315	8,268	C:86.8 % [S:86.2 %, D:0.6 %], F:7.0 %, M:6.2 %, n:1,315	NCBI
Xanthoria parietina	XPAR	31.9 Mb	1,731,186	39	302	C:95.2 % [S:95.1 %, D:0.1 %], F:3.4 %, M:1.4 %, n:1,315	11,065	C:95.4 % [S:95.2 %, D:0.2 %], F:3.5 %, M:1.1 %, n:1,315	JGI

Table S2. Overview of Biosynthetic Gene Clusters in *Evernia prunastri*.

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
1	epruFC2	175925	197657	Terpene	EPRU-FC_00062		terpene synthase metal binding domain protein	-	-
2	epruFC2	435175	499921	R-PKS	EPRU-FC_00156	KS-AT-DH-cMT-ER-KR-ACP	Acyl transferase/acyl hydrolase/lysophospholipase	-	-
3	epruFC4	104371	189977	PKS-NRPS	EPRU-FC_00468	KS-AT-DH-cMT-KR-C-A-PCP-TD	hybrid PKS/NRPS enzyme, putative	Aspyridone biosynthetic gene cluster (33 % of genes show similarity)	BGC0000959_c1
4	epruFC5	219484	260867	T3PKS	EPRU-FC_01444		chalcone synthase	-	-
5	epruFC5	539898	570865	R-PKS	EPRU-FC_01523	KS-AT-DH-ACP-cMT	polyketide synthase	-	-
6	epruFC6	276044	297335	Terpene	EPRU-FC_01153		terpene synthase metal binding domain protein	-	-
7	epruFC9	441924	463735	Terpene	EPRU-FC_01357		Terpene synthase metal binding domain protein	-	-
8	epruFC9	468189	516248	PKS-NRPS	EPRU-FC_01369	KS-AT-DH-cMT-KR-C-A-PCP-TD	polyketide synthetase	-	-
9	epruFC11	223866	271871	NR-PKS	EPRU-FC_01722	KS-AT-ACP-cMT	polyketide synthase	-	-
10	epruFC11	445831	488836	R-PKS	EPRU-FC_01787	KS-AT-DH-ER-KR	polyketide synthase, putative	-	-
11	epruFC18	21	48283	R-PKS	EPRU-FC_02335	KS-AT-DH-cMT-ER-KR-ACP	polyketide synthase	-	-
12	epruFC19	282678	303953	Terpene	EPRU-FC_02533		Terpenoid synthase	-	-
13	epruFC19	358412	402330	-	-		-	-	-
14	epruFC20	21012	67873	R-PKS	EPRU-FC_02707	KS-AT-DH-ACP	polyketide synthase	-	-
15	epruFC20	121918	143577	Terpene	EPRU-FC_02735		farnesyl-diphosphate farnesyltransferase	-	-
16	epruFC21	176044	223963	NR-PKS	EPRU-FC_02873	KS-AT-ACP-cMT	polyketide synthase	-	-
17	epruFC21	253660	309261	-	-		-	-	-
18	epruFC23	80769	124252	-	-		-	-	-
19	epruFC23	157176	178560	Indole	EPRU-FC_03402		dimethylallyl tryptophan synthase	-	-
20	epruFC23	215104	262085	R-PKS	EPRU-FC_03419	KS-AT-DH-cMT-ER_KR-ACP	polyketide synthase, putative	Byssochlamic acid biosynthetic gene cluster (20 % of genes show similarity)	BGC0001340_c1
21	epruFC24	68703	111618	Other	EPRU-FC_03993	A-TD	nonribosomal peptide synthase, putative	-	-
22	epruFC25	160663	185304	Terpene	EPRU-FC_04916		squalene-hopene cyclase	-	-

– continued on next page

Table S2 – *continued from previous page*

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
23	epruFC26	1	13755	Terpene	EPRU-FC_03461		Lycopene beta-cyclase	-	-
24	epruFC28	182797	225667	Other	EPRU-FC_03629	A-TE	NRPS-like enzyme, putative	-	-
25	epruFC31	1	44154	R-PKS	EPRU-FC_04169	KS-AT-DH-ER-KR-ACP	related to polyketide synthase	-	-
26	epruFC33	129973	173183	Other	EPRU-FC_03720	A-TD	NRPS-like enzyme	-	-
27	epruFC33	219614	241181	Terpene	EPRU-FC_03746		Terpenoid synthase	-	-
28	epruFC36	221797	265140	Other	EPRU-FC_03003	A-TD	NRPS-like enzyme	-	-
29	epruFC37	73201	123800	R-PKS	EPRU-FC_05147	KS-AT-DH-ACP-cMT	BcPKS16, polyketide synthase	-	-
30	epruFC39	103379	151675	R-PKS	EPRU-FC_05073	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
31	epruFC44	144296	187746	NRPS	EPRU-FC_06064	A-PCP-C	Nonribosomal Peptide Synthase (NRPS)	-	-
32	epruFC46	253755	289622	R-PKS	EPRU-FC_04343	AT_KS-AT-DH-ACP-TE	Acyl transferase/acyl hydrolase/lysophospholipase	-	-
33	epruFC47	1	37200	R-PKS	EPRU-FC_04788	KS_AT-DH-ACP-TE	polyketide synthase, putative	-	-
34	epruFC47	67091	78486	Siderophore	EPRU-FC_04804		predicted protein	-	-
35	epruFC48	89638	136397	NR-PKS	EPRU-FC_05300	KS-AT-ACP-ACP-TE	polyketide synthase	Grayanic acid biosynthetic gene cluster (66 % of genes show similarity)	BGC0001266_c1
36	epruFC50	207505	255685	R-PKS	EPRU-FC_05262	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
37	epruFC55	1	41348	PKS-NRPS	EPRU-FC_07216	KS-AT-DH-cMT-KR-ACP-C-A-PCP-TD	Hybrid PKS-NRPS	-	-
38	epruFC55	45219	109079	NRPS	EPRU-FC_07229	A-PCP-E-C-A-PCP-C-A-PCP-E-C-A-PCP-C-A-PCP-E-C-PCP-C	nonribosomal peptide synthase	-	-
39	epruFC55	140084	183415	Other	EPRU-FC_07252	A-TD	NRPS-like enzyme	-	-
40	epruFC59	86584	130658	Other	EPRU-FC_06855	A-PCP-NAD	alpha-aminoadipate reductase large subunit, putative	-	-
41	epruFC61	107684	154305	NR-PKS	EPRU-FC_07415	KS-AT-ACP-cMT	polyketide synthase	-	-
42	epruFC63	1	27241	R-PKS	EPRU-FC_04960	KS-AT-DH-KR-ACP	6-methylsalicylic acid synthase	-	-
43	epruFC64	127999	171071	NR-PKS	EPRU-FC_06328	KS-AT	putative polyketide synthase	-	-
44	epruFC65	101688	145560	Other	EPRU-FC_05810	A-TD	putative NRPS-like enzyme	-	-
45	epruFC66	113022	135361	Terpene	EPRU-FC_07778		oxidosqualene:lanosterol cyclase	-	-
46	epruFC66	172410	216643	R-PKS	EPRU-FC_07791	KS-AT-DH-ER-KR-ACP	putative polyketide synthase protein	Brefeldin biosynthetic gene cluster (20 % of genes show similarity)	BGC0001141_c1

– *continued on next page*

Table S2 – *continued from previous page*

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
47	epruFC69	21857	88310	NRPS	EPRU-FC_07598	A-PCP	nonribosomal peptide synthase, putative	-	-
				Indole	EPRU-FC_07608		aromatic prenyltransferase		
48	epruFC69	164797	208103	Other	EPRU-FC_07637	A-PCP-NAD	NRPS-like enzyme	-	-
49	epruFC71	86813	129929	Other	EPRU-FC_06798	A-TD	non-ribosomal peptide synthetase	-	-
50	epruFC82	1	42075	R-PKS	EPRU-FC_06660	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
51	epruFC84	124822	181022	NR-PKS	EPRU-FC_08264	KS-AT-ACP-TE	polyketide synthase	-	-
				R-PKS	EPRU-FC_08265	KS-AT-DH-ER-KR	Beta-ketoacyl synthase		
52	epruFC87	122993	170830	NR-PKS	EPRU-FC_07477	KS-AT-ACP-cMT-TD	Male sterility, NAD-binding	Citrinin biosynthetic gene cluster (56 % of genes show similarity)	BGC0001338_c1
53	epruFC91	88257	133867	NR-PKS	EPRU-FC_08170	KS-AT-ACP	ketoacyl-synt-domain-containing protein	Emericellin biosynthetic gene cluster (42 % of genes show similarity)	BGC0001271_c1
54	epruFC91	139206	171511	Other	EPRU-FC_08180	A-ACP-TD	nonribosomal peptide synthetase	-	-
55	epruFC92	1971	44922	Other	EPRU-FC_07055	A-ACP-TE	TdiA protein	-	-
56	epruFC92	84761	128409	Other	EPRU-FC_07078	A-TD	thioester reductase domain-containing protein	-	-
57	epruFC93	1	31996	R-PKS	EPRU-FC_07799	KS-AT-DH-ER-KR-ACP	polyketide synthase	-	-
58	epruFC94	129486	150724	Terpene	EPRU-FC_07677		Squalene/phytoene synthase	-	-
59	epruFC95	38525	89193	PKS-NRPS	EPRU-FC_08659	KS-AT-DH-ER-KR-C-A-PCP	Hybrid PKS-NRPS	-	-
60	epruFC97	62792	106078	-	EPRU-FC_08113	KS_KS-AT-DH_ER-KR	polyketide synthase	-	-
61	epruFC99	91200	144325	R-PKS	EPRU-FC_08221	KS-AT-DH-TE	polyketide synthase	-	-
				Other	EPRU-FC_08222	A-TD	NRPS-like enzyme, putative		
62	epruFC110	1	33705	R-PKS	EPRU-FC_08925	KS-AT-DH-ER-KR	polyketide synthase	-	-
63	epruFC111	94582	132781	NRPS	EPRU-FC_09130	A-C-A-PCP-C-PCP-C-C	nonribosomal siderophore peptide synthase SidC	-	-
64	epruFC112	36283	84562	R-PKS	EPRU-FC_09516	KS-AT-DH-cMT-ER-KR	polyketide synthase	-	-
65	epruFC115	1	24855	Other	EPRU-FC_09385	ACP-TD	acetyl-CoA synthetase-like protein	-	-
66	epruFC120	55131	106632	R-PKS	EPRU-FC_09662	KS_AT-DH-cMT-ER-KR	Putative polyketide synthase	-	-
				T3PKS	EPRU-FC_09663		chalcone synthase B		
67	epruFC126	71231	107400	-	-		-	-	-

– continued on next page

Table S2 – *continued from previous page*

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
68	epruFC128	2183	50481	R-PKS	EPRU-FC_10011	KS-AT-DH-ACP-cMT	related to polyketide synthase	Shanorellin biosynthetic gene cluster (28 % of genes show similarity)	BGC0001219_c1
69	epruFC131	1	33479	NR-PKS	EPRU-FC_09326	KS-AT-ACP-ACP-TE	BcPKS12, polyketide synthase	-	-
70	epruFC135	23450	45625	Terpene	EPRU-FC_09953		squalene cyclase	-	-
71	epruFC138	10525	56041	R-PKS	EPRU-FC_09573	KS-AT-DH-ACP-TE	Type I Iterative Polyketide synthase (PKS)	-	-
72	epruFC150	1	46610	R-PKS	EPRU-FC_10219	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
73	epruFC157	8432	55118	R-PKS	EPRU-FC_09854	KS-AT-DH-ACP-ACP-TE	conidial yellow pigment biosynthesis polyketide synthase	-	-
74	epruFC165	1	68454	R-PKS	EPRU-FC_10046	KS-AT-DH-cMT-ER-KR-ACP	Acyl transferase/acyl hydrolase/lysophospholipase	-	-
74				R-PKS	EPRU-FC_10052	KS-AT-DH-cMT-ER-KR	polyketide synthase, putative		
75	epruFC178	22953	60149	R-PKS	EPRU-FC_10576	KS-AT-DH-cMT-ER-KR-ACP	Type I Iterative Polyketide synthase (PKS)	-	-
76	epruFC193	19900	53880	R-PKS	EPRU-FC_10360	KS-AT-DH-ER-KR-ACP	related to polyketide synthase	-	-
77	epruFC203	5019	26395	Terpene	EPRU-FC_10616		squalene cyclase	-	-
78	epruFC206	7352	41211	PKS	EPRU-FC_10661	KS	polyketide synthase	-	-
				NRPS	EPRU-FC_10659	C-A-PCP-C	BcNRPS1, nonribosomal peptide synthetase		
79	epruFC226	1	15080	Terpene	EPRU-FC_10734		probable farnesyltranstransferase (al-3)	-	-
80	epruFC253	1	11353	-	-		-	-	-

Table S3. Overview of Biosynthetic Gene Clusters in *Pseudevernia furfuracea*.

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
Cluster	Scaffold	From	To	Cluster type	Gene	Domains/Notes	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
1	pfurFC1	562783	610207	R-PKS	PFUR-FC_00138	KS-AT-DH-ER-KR	reducing type I polyketide synthase 10	-	-
2	pfurFC1	1121302	1173255	PKS-NRPS	PFUR-FC_00284	KS-AT-DH-cMT-KR-C-A-PCP-TD	lovastatin nonaketide synthase	Aspyridone biosynthetic gene cluster (22 % of genes show similarity)	BGC0000959_c1
3	pfurFC1	1275530	1332188	NR-PKS	PFUR-FC_00325	KS-AT-ACP-TE	polyketide synthase	-	-
				R-PKS	PFUR-FC_00326	KS-AT-DH-ER-KR-ACP	Beta-ketoacyl synthase		
4	pfurFC2	121832	169239	R-PKS	PFUR-FC_01387	KS-AT-DH-ER-KR-ACP	polyketide synthase	-	-
5	pfurFC2	408152	456230	R-PKS	PFUR-FC_01455	KS-AT-DH-ER-KR	polyketide synthase-nonribosomal peptide synthetase	-	-
6	pfurFC2	535308	556823	Indole	PFUR-FC_01480		dimethylallyl tryptophan synthase	-	-
7	pfurFC2	1068180	1117253	R-PKS	PFUR-FC_01616	KS-AT-DH-cMT-KR-TD	polyketide synthase	Betaenone C / betaenone A biosynthetic gene cluster (25 % of genes show similarity)	BGC0001280_c1
8	pfurFC2	2124717	2146373	Terpene	PFUR-FC_01880		farnesyl-diphosphate farnesyltransferase	-	-
9	pfurFC3	19588	63188	Other	PFUR-FC_00759	A-NAD	L-aminoadipate-semialdehyde dehydrogenase	-	-
10	pfurFC3	2020441	2092163	R-PKS	PFUR-FC_01263	KS-AT-DH-ER-KR	beta-ketoacyl synthase domain-containing protein	Brefeldin biosynthetic gene cluster (20 % of genes show similarity)	BGC0001141_c1
				Other	PFUR-FC_01255	A-TD	putative NRPS-like enzyme		
11	pfurFC3	2415629	2459083	Other	PFUR-FC_01357	A-TD	thioester reductase domain-containing protein	-	-
12	pfurFC4	32599	83599	NRPS	PFUR-FC_01993	A-PCP-C-A-C-PCP-C-C	nonribosomal siderophore peptide synthase SidC	-	-

– continued on next page

Table S3 – *continued from previous page*

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
13	pfurFC4	162468	210590	R-PKS	PFUR-FC_02025	KS-AT-DH-cMT-ER-KR	polyketide synthase	Solanapyrone biosynthetic gene cluster (33 % of genes show similarity)	BGC0000146_c1
14	pfurFC4	596479	643127	R-PKS	PFUR-FC_02144	KS-AT-DH-ACP-cMT	BcPKS16, polyketide synthase	-	-
15	pfurFC4	1028691	1077227	R-PKS	PFUR-FC_02262	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
16	pfurFC4	1274798	1295885	Terpene	PFUR-FC_02280		Aristolochene synthase	-	-
17	pfurFC4	1322625	1368918	NR-PKS	PFUR-FC_02294	KS-AT-ACP-ACP-TE	Acyl transferase/acyl hydrolase/lysophospholipase	Grayanic acid biosynthetic gene cluster (66 % of genes show similarity)	BGC0001266_c1
18	pfurFC5	297002	348409	R-PKS	PFUR-FC_03717	KS-AT-DH-cMT-ER-KR	putative polyketide synthase	-	-
				T3PKS	PFUR-FC_03718		chalcone synthase B		
19	pfurFC5	1214592	1261281	R-PKS	PFUR-FC_03906	KS-AT-DH-ACP-ACP-TE	polyketide synthase	Terrein biosynthetic gene cluster (36 % of genes show similarity)	BGC0000161_c1
20	pfurFC5	1460690	1507114	NRPS	PFUR-FC_03956	A-PCP-C-A-C	nonribosomal peptide synthase SidE	-	-
21	pfurFC7	1000723	1043755	-	-		-	-	-
22	pfurFC7	1610413	1633037	Terpene	PFUR-FC_03648		squalene cyclase	-	-
23	pfurFC8	237447	299968	R-PKS	PFUR-FC_02504	KS-AT-DH-ER-KR	Type I Iterative Polyketide synthase (PKS)	-	-
				Other	PFUR-FC_02502	A-TD	non-ribosomal peptide synthetase		
24	pfurFC9	6722	50245	Other	PFUR-FC_04327	A-TD	NRPS-like enzyme	-	-
25	pfurFC9	101616	145497	Other	PFUR-FC_04353	A-PCP-NAD	L-aminoadipate-semialdehyde dehydrogenase large subunit	-	-
26	pfurFC10	792474	814434	Terpene	PFUR-FC_05461		Lycopene beta-cyclase	-	-
27	pfurFC11	165658	209059	Other	PFUR-FC_04042	A-TD	acetyl-CoA synthetase-like protein	-	-
28	pfurFC12	805117	858669	R-PKS	PFUR-FC_04901	KS-AT-ACP-ACP-TE-KR	polyketide synthase	-	-
29	pfurFC13	1077300	1098721	Indole	PFUR-FC_05263		dimethylallyl tryptophan synthase	-	-
30	pfurFC15	387999	439243	R-PKS	PFUR-FC_05902	KS-AT-DH-cMT	polyketide synthase	-	-
				Other	PFUR-FC_05903	A-PCP-TD	NRPS-like enzyme		

– *continued on next page*

Table S3 – *continued from previous page*

Cluster	Scaffold	From	To	Cluster type	Core gene	Domains	Blast2GO description	Most similar known cluster	MIBiG BGC-ID
31	pfurFC16	294526	341886	R-PKS	PFUR-FC_06732	KS_AT-DH-cMT-ER-KR	polyketide synthase, putative	Byssochlamic acid biosynthetic gene cluster (25 % of genes show similarity)	BGC0001340_c1
32	pfurFC16	446622	467785	Terpene	PFUR-FC_06764		geranylgeranyl pyrophosphate synthase	-	-
33	pfurFC16	482462	530388	NR-PKS	PFUR-FC_06778	KS-AT-ACP-cMT-TD	polyketide synthase, putative	-	-
34	pfurFC17	367411	410336	Other	PFUR-FC_06983	A-PCP-TE	putative non-ribosomal peptide synthase-like protein	-	-
35	pfurFC18	332561	381013	R-PKS	PFUR-FC_07626	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
36	pfurFC18	763979	807436	NRPS	PFUR-FC_07703	A-PCP-C	peptide synthetase	-	-
37	pfurFC20	94882	141599	NR-PKS	PFUR-FC_07085	KS-AT-ACP-ACP-TE	BcPKS12, polyketide synthase	-	-
38	pfurFC21	504683	548936	Other	PFUR-FC_06211	A-TD	nonribosomal peptide synthase, putative	-	-
39	pfurFC23	454460	500051	NR-PKS	PFUR-FC_06409	KS-AT-ACP	polyketide synthase	Pestheic acid biosynthetic gene cluster (15 % of genes show similarity)	BGC0000121_c1
40	pfurFC23	603120	671736	T3PKS	PFUR-FC_06452		chalcone synthase	-	-
41	pfurFC23	674459	713509	R-PKS	PFUR-FC_06464	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
42	pfurFC24	352574	400864	R-PKS	PFUR-FC_08419	KS-AT-DH-cMT-ER-KR-ACP	putative polyketide synthase	-	-
43	pfurFC25	265354	308560	Other	PFUR-FC_07480	A-TD	NRPS-like enzyme	-	-
44	pfurFC25	428943	471813	Other	PFUR-FC_07518	A-TE	TdiA protein	-	-
45	pfurFC26	126642	215199	PKS-NRPS	PFUR-FC_07753	KS-AT-DH-cMT-KR-ACP-C-A-PCP-TD	Hybrid PKS-NRPS	-	-
				NRPS	PFUR-FC_07750	A-PCP-E-C-A-PCP-C-A-PCP-E-C-A-PCP-C-A-PCP-E-C-PCP-C	nonribosomal peptide synthase		
46	pfurFC27	111648	186892	R-PKS	PFUR-FC_08106	KS-AT-DH-ER-KR	related to polyketide synthase	-	-
47	pfurFC28	359300	407656	R-PKS	PFUR-FC_08064	KS-AT-DH-cMT-ER-KR-ACP	Lovastatin diketide synthase LovF	-	-
48	pfurFC30	8866	59491	PKS-NRPS	PFUR-FC_08252	KS-AT-DH-ER-KR-C-A-PCP	Hybrid PKS-NRPS	-	-
49	pfurFC36	10296	57325	R-PKS	PFUR-FC_08450	KS-AT-DH-KR_ACP	polyketide synthase	-	-
50	pfurFC37	53532	97152	Other	PFUR-FC_08687	A-TD	thioester reductase domain-containing protein	-	-
51	pfurFC38	94358	160893	R-PKS	PFUR-FC_08573	KS-AT_DH-ER-KR-ACP	polyketide synthase	-	-

Table S6. Reciprocal Best Blast Hits of *Evernia prunastri* and *Pseudevernia furfuracea* Cluster Genes

***Evernia prunastri* cluster genes**	***Pseudevernia furfuracea* cluster genes**
EPRU-FC_01151	PFUR-FC_08110
EPRU-FC_01435	PFUR-FC_06462
EPRU-FC_01438	PFUR-FC_06461
EPRU-FC_01442	PFUR-FC_06460
EPRU-FC_01444	PFUR-FC_06452
EPRU-FC_01445	PFUR-FC_06451
EPRU-FC_01789	PFUR-FC_08449
EPRU-FC_02703	PFUR-FC_00760
EPRU-FC_02732	PFUR-FC_01877
EPRU-FC_02734	PFUR-FC_01879
EPRU-FC_02735	PFUR-FC_01880
EPRU-FC_02737	PFUR-FC_01882
EPRU-FC_02897	PFUR-FC_03498
EPRU-FC_03382	PFUR-FC_02506
EPRU-FC_03386	PFUR-FC_02507
EPRU-FC_03387	PFUR-FC_02508
EPRU-FC_03388	PFUR-FC_02509
EPRU-FC_03389	PFUR-FC_02510
EPRU-FC_03402	PFUR-FC_05263
EPRU-FC_03461	PFUR-FC_05461
EPRU-FC_03462	PFUR-FC_05460
EPRU-FC_03463	PFUR-FC_05459
EPRU-FC_03623	PFUR-FC_07514
EPRU-FC_03624	PFUR-FC_07515
EPRU-FC_03626	PFUR-FC_07522
EPRU-FC_03627	PFUR-FC_07521
EPRU-FC_03628	PFUR-FC_07519
EPRU-FC_03629	PFUR-FC_07518
EPRU-FC_05143	PFUR-FC_02140
EPRU-FC_05146	PFUR-FC_02141
EPRU-FC_05147	PFUR-FC_02144
EPRU-FC_05149	PFUR-FC_02145
EPRU-FC_05150	PFUR-FC_02146
EPRU-FC_05151	PFUR-FC_02147
EPRU-FC_05297	PFUR-FC_02298
EPRU-FC_05299	PFUR-FC_02295
EPRU-FC_05300	PFUR-FC_02294
EPRU-FC_05301	PFUR-FC_02293
EPRU-FC_05804	PFUR-FC_01248
EPRU-FC_05806	PFUR-FC_01250
EPRU-FC_05808	PFUR-FC_01252
EPRU-FC_05810	PFUR-FC_01255
EPRU-FC_05812	PFUR-FC_01264
EPRU-FC_06063	PFUR-FC_07706
EPRU-FC_06065	PFUR-FC_07704
EPRU-FC_06068	PFUR-FC_07698
EPRU-FC_06069	PFUR-FC_07699
EPRU-FC_06797	PFUR-FC_02501
EPRU-FC_06798	PFUR-FC_02502
EPRU-FC_06849	PFUR-FC_04350
EPRU-FC_06853	PFUR-FC_04351
EPRU-FC_06854	PFUR-FC_04352
EPRU-FC_06855	PFUR-FC_04353
EPRU-FC_06859	PFUR-FC_04354
EPRU-FC_06861	PFUR-FC_04358
EPRU-FC_07073	PFUR-FC_08690
EPRU-FC_07074	PFUR-FC_08689
EPRU-FC_07075	PFUR-FC_08688
EPRU-FC_07078	PFUR-FC_08687
EPRU-FC_07079	PFUR-FC_08686
EPRU-FC_07080	PFUR-FC_08685
EPRU-FC_07082	PFUR-FC_08683
EPRU-FC_07216	PFUR-FC_07753
EPRU-FC_07217	PFUR-FC_07754
EPRU-FC_07219	PFUR-FC_07755
EPRU-FC_07221	PFUR-FC_07757
EPRU-FC_07222	PFUR-FC_07758
EPRU-FC_07223	PFUR-FC_07759
EPRU-FC_07224	PFUR-FC_07760
EPRU-FC_07227	PFUR-FC_07752
EPRU-FC_07228	PFUR-FC_07751
EPRU-FC_07229	PFUR-FC_07750
EPRU-FC_07232	PFUR-FC_07749
EPRU-FC_07233	PFUR-FC_07748
EPRU-FC_07234	PFUR-FC_07747
EPRU-FC_07609	PFUR-FC_01479
EPRU-FC_07638	PFUR-FC_04897
EPRU-FC_08106	PFUR-FC_08584
EPRU-FC_08107	PFUR-FC_08583
EPRU-FC_08108	PFUR-FC_08581
EPRU-FC_08110	PFUR-FC_08576
EPRU-FC_08112	PFUR-FC_08573
EPRU-FC_08114	PFUR-FC_08572
EPRU-FC_08115	PFUR-FC_08570
EPRU-FC_08117	PFUR-FC_08567
EPRU-FC_08118	PFUR-FC_08566
EPRU-FC_08119	PFUR-FC_08565
EPRU-FC_08120	PFUR-FC_08564
EPRU-FC_08162	PFUR-FC_02028
EPRU-FC_08177	PFUR-FC_04326
EPRU-FC_08180	PFUR-FC_04327
EPRU-FC_08181	PFUR-FC_04328

– continued on next page

Table S6 – *continued from previous page*

***Evernia prunastri* cluster genes**	***Pseudevernia furfuracea* cluster genes**
EPRU-FC_08264	PFUR-FC_00325
EPRU-FC_08265	PFUR-FC_00326
EPRU-FC_08656	PFUR-FC_08254
EPRU-FC_08657	PFUR-FC_08253
EPRU-FC_08659	PFUR-FC_08252
EPRU-FC_09124	PFUR-FC_01986
EPRU-FC_09125	PFUR-FC_01987
EPRU-FC_09126	PFUR-FC_01988
EPRU-FC_09127	PFUR-FC_01989
EPRU-FC_09128	PFUR-FC_01990
EPRU-FC_09129	PFUR-FC_01992
EPRU-FC_09130	PFUR-FC_01993
EPRU-FC_09325	PFUR-FC_07084
EPRU-FC_09326	PFUR-FC_07085
EPRU-FC_09327	PFUR-FC_07086
EPRU-FC_09328	PFUR-FC_07087
EPRU-FC_09386	PFUR-FC_04041
EPRU-FC_09656	PFUR-FC_03713
EPRU-FC_09657	PFUR-FC_03714
EPRU-FC_09658	PFUR-FC_03715
EPRU-FC_09659	PFUR-FC_03716
EPRU-FC_09661	PFUR-FC_03717
EPRU-FC_09663	PFUR-FC_03718
EPRU-FC_09664	PFUR-FC_03719
EPRU-FC_09952	PFUR-FC_03647
EPRU-FC_09953	PFUR-FC_03648
EPRU-FC_09954	PFUR-FC_03649
EPRU-FC_09956	PFUR-FC_03651
EPRU-FC_10219	PFUR-FC_07626
EPRU-FC_10360	PFUR-FC_08106
EPRU-FC_10734	PFUR-FC_06764
EPRU-FC_10735	PFUR-FC_06765
EPRU-FC_10736	PFUR-FC_06766
EPRU-FC_10737	PFUR-FC_06767

Table S7. List of Putative Regulatory Genes in *Evernia prunastri* (EPRU) and *Pseudevernia furfuracea* (PFUR)

Cluster	Gene name	Blast2GO description
EPRU01	EPRU-FC_00065	predicted protein
EPRU02	EPRU-FC_00144	Ankyrin repeat-containing protein
EPRU02	EPRU-FC_00152	Zn2/Cys6 DNA-binding protein
EPRU02	EPRU-FC_00159	phosphatidyl inositol-specific phospholipase C
EPRU02	EPRU-FC_00161	zinc finger transcription factor ace1
EPRU05	EPRU-FC_01518	related to nitrate assimilation regulatory protein nirA
EPRU05	EPRU-FC_01522	C6 zinc finger domain protein
EPRU13	EPRU-FC_02558	related to regulatory protein for the arginine catabolic pathway
EPRU14	EPRU-FC_02704	C6 finger domain protein GliZ-like, putative
EPRU14	EPRU-FC_02706	Zn2/Cys6 DNA-binding protein
EPRU14	EPRU-FC_02708	GntR family transcriptional regulator
EPRU15	EPRU-FC_02730	RNA polymerase sigma-70 factor, ECF subfamily
EPRU16	EPRU-FC_02871	C6 transcription factor (AflR)
EPRU17	EPRU-FC_02904	C6 zinc finger domain protein
EPRU20	EPRU-FC_03414	predicted glutathione s-transferase
EPRU20	EPRU-FC_03428	fungal specific transcription factor domain-containing protein
EPRU29	EPRU-FC_05154	C6 zinc finger domain protein
EPRU31	EPRU-FC_06067	C6 zinc finger domain-containing protein
EPRU37	EPRU-FC_07222	predicted protein
EPRU40	EPRU-FC_06848	C6 zinc finger domain-containing protein
EPRU40	EPRU-FC_06857	sigma-54-dependent Fis family transcriptional regulator
EPRU41	EPRU-FC_07411	related to Cutinase transcription factor 1 alpha
EPRU42	EPRU-FC_04961	fungal specific transcription factor domain-containing protein
EPRU42	EPRU-FC_04962	fungal specific transcription factor domain-containing protein
EPRU46	EPRU-FC_07792	tetratricopeptide ankyrin repeat and coiled-coil containing 1
EPRU46	EPRU-FC_07794	Fungal specific transcription factor domain-containing protein
EPRU49	EPRU-FC_06796	choline kinase
EPRU50	EPRU-FC_06663	Zn2/Cys6 DNA-binding protein
EPRU50	EPRU-FC_06664	chromosome segregation protein SMC
EPRU52	EPRU-FC_07480	citrinin biosynthesis transcriptional activator CtnR
EPRU54	EPRU-FC_08182	C6 zinc finger domain protein
EPRU54	EPRU-FC_08183	hypothetical protein ST47_g2692
EPRU56	EPRU-FC_07083	vacuolar protein 8
EPRU59	EPRU-FC_08658	protein kinase subdomain-containing protein
EPRU67	EPRU-FC_09801	Zn2/Cys6 DNA-binding protein
EPRU68	EPRU-FC_10005	Ca2+-binding protein, RTX toxin-related
EPRU68	EPRU-FC_10008	Fungal Zn2-Cys6 binuclear cluster domain-containing protein
EPRU69	EPRU-FC_09327	transcription factor
EPRU69	EPRU-FC_09331	HET-domain-containing protein
EPRU71	EPRU-FC_09575	RNA polymerase sigma-70 factor, ECF subfamily
EPRU71	EPRU-FC_09578	Putative DNA-binding protein ESCAROLA
EPRU73	EPRU-FC_09850	hypothetical protein V499_04723
EPRU73	EPRU-FC_09855	myb transcription protein
EPRU77	EPRU-FC_10613	C6 transcription factor
PFUR01	PFUR-FC_00133	c6 finger domain-containing protein
PFUR07	PFUR-FC_01611	Putative transcriptional regulatory protein C15D4.02
PFUR07	PFUR-FC_01619	predicted protein

– continued on next page

Table S7 – *continued from previous page*

Cluster	Gene name	Blast2GO description
PFUR11	PFUR-FC_01356	transposase tc1-like protein
PFUR14	PFUR-FC_02149	Transcription factor, fungi
PFUR20	PFUR-FC_03954	transcription factor AbaA
PFUR23	PFUR-FC_02500	choline kinase
PFUR24	PFUR-FC_04329	Fungal Zn binuclear cluster domain containing protein
PFUR31	PFUR-FC_06727	ketosteroid isomerase-like protein
PFUR35	PFUR-FC_07627	sensor histidine kinase
PFUR36	PFUR-FC_07702	Zn2/Cys6 DNA-binding protein
PFUR36	PFUR-FC_07708	BTB/POZ domain-containing protein KCTD3
PFUR37	PFUR-FC_07086	transcription factor
PFUR39	PFUR-FC_06405	Aflatoxin biosynthesis regulatory protein
PFUR39	PFUR-FC_06406	Aflatoxin regulatory protein
PFUR45	PFUR-FC_07758	predicted protein

Figure S1. Phylogenetic tree based on the ketosynthase (KS) domain.

RAxML tree based on
- the protein sequence of the ketosynthase (KS) domain
- 550 entries (413 sequences from lichen-forming fungi, 131 characterized metabolites from MIBiG, 6 animal fatty acid synthases)

MIBiG (Minimum Information about a Biosynthetic Gene cluster) entries with characterized main products

BGC0001205 communesin KS-AT-DH-ER-KR *Penicillium expansum*

MIBiG accession | Main product | Domain annotation | Organism

Entries from lichen-forming fungi

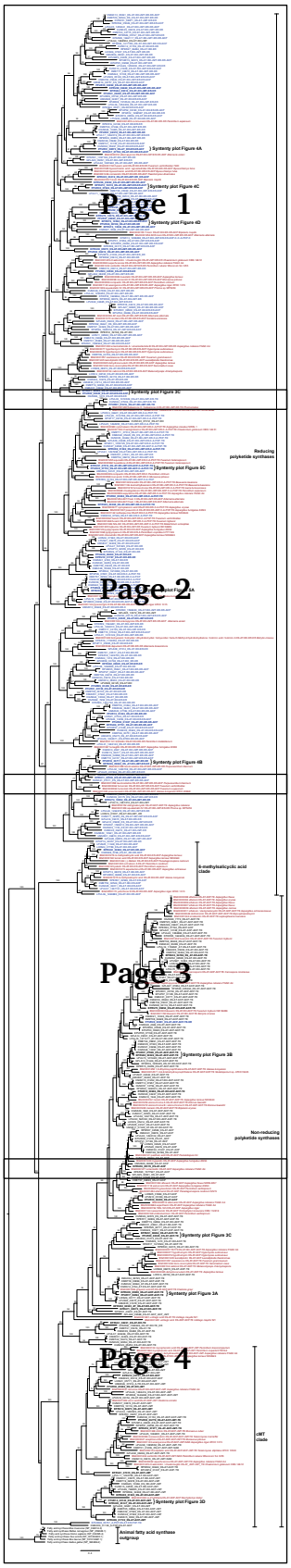
PFUR37_07085_KS-AT-ACP-ACP-TE

Species (see table for abbreviation) | Cluster number | Gene name number | Domain annotation

The two focus species *Evernia prunastri* and *Pseudevernia furfuracea* are highlighted in **bold**. Blues entries indicate a domain annotation with ER or KR domains.

Abbreviations

CGRA = *Cladonia grayi*	A = Adenylation
CMAC = *Cladonia macilenta*	ACP = Acyl carrier protein
CMET = *Cladonia metacorallifera*	AT = Acyltransferase
CUNC = *Cladonia uncialis*	C = Condensation
EPUP = *Endocarpon pusillum* (Park et al.)	cMT = C-Methyltransferase
EPUW = *Endocarpon pusillum* (Wang et al.)	DH = Dehydratase
EPRU = ***Evernia prunastri***	E = Epimerization
GFLA = *Gyalolechia flavorubescens*	ER = Enoylreductase
LPUL = *Lobaria pulmonaria*	KR = Ketoreductase
PFUR = ***Pseudevernia furfuracea***	KS = Ketosynthase
RPER = *Ramalina peruviana*	PCP = Peptide carrier protein
LHIS = *Umbilicaria hispanica*	T = Thiolation
UMUE = *Umbilicaria muehlenbergii*	TD = Terminal domain
LPUS = *Umbilicaria pustulata*	TE = Thioesterase
XPAR = *Xanthoria parietina*	



The phylogenetic tree is presented in detail on the following four separate pages.

Figure S1. – *continued (Detail page 1 of 4)*

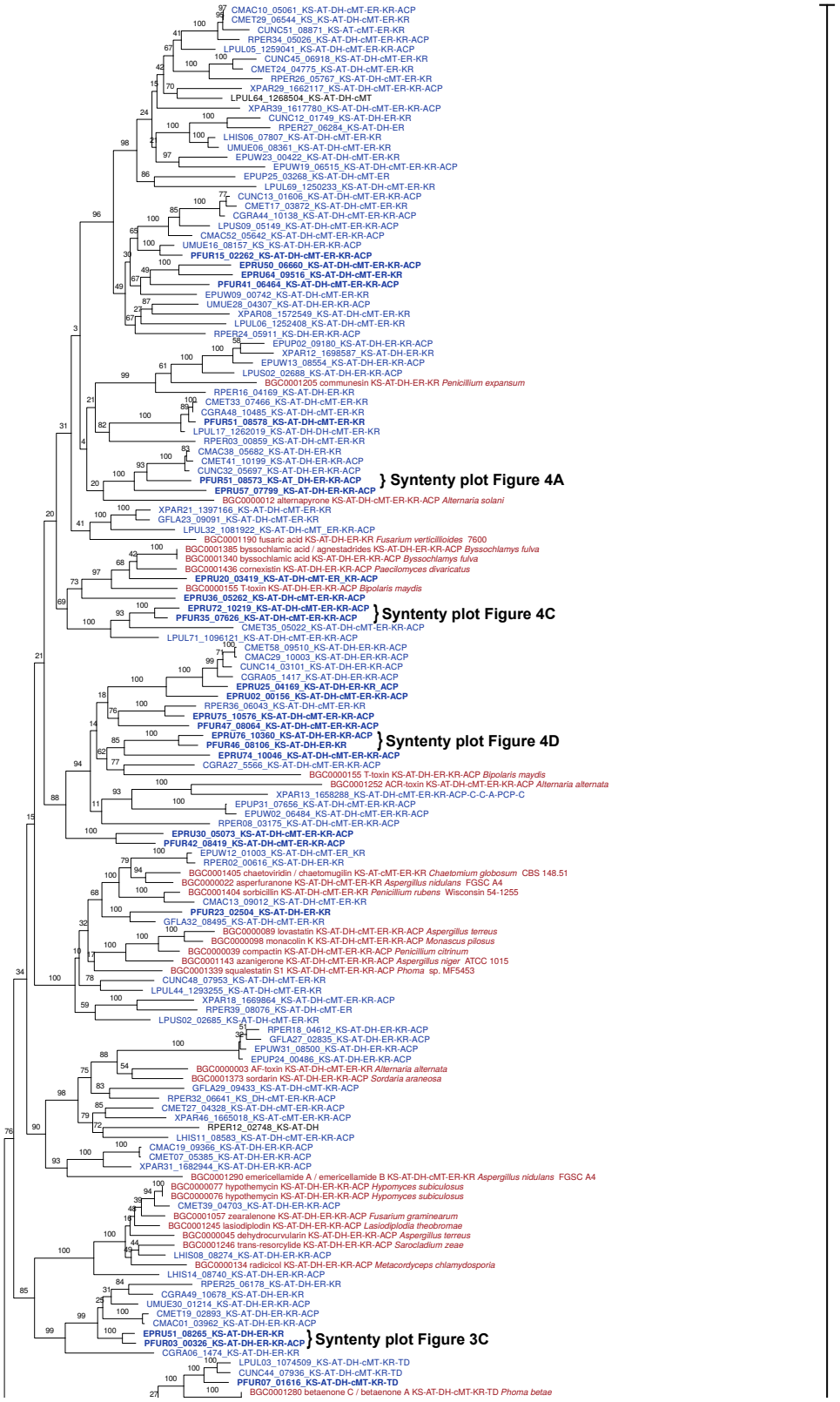


Figure S1. – *continued (Detail page 2 of 4)*

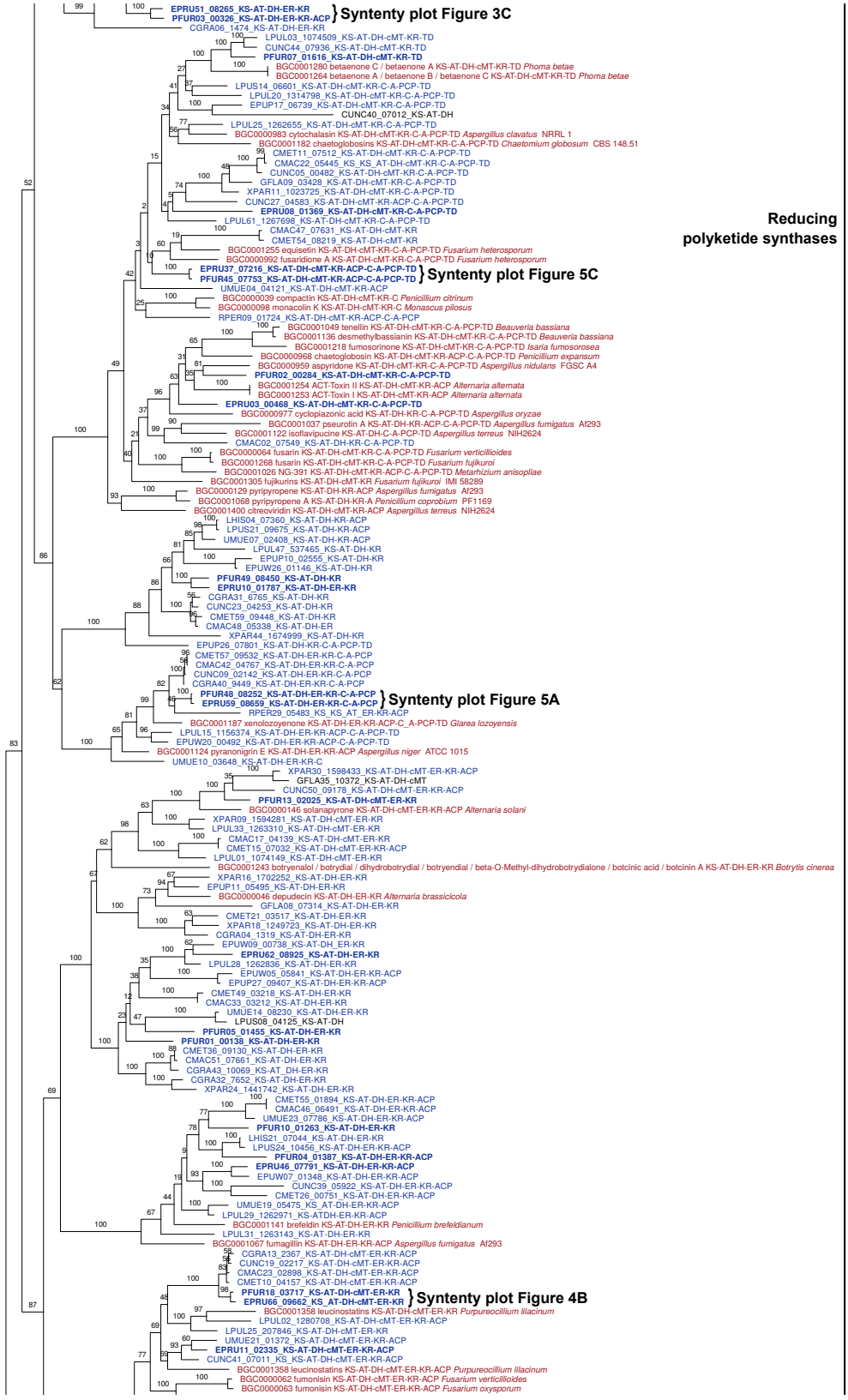


Figure S1. – *continued (Detail page 3 of 4)*

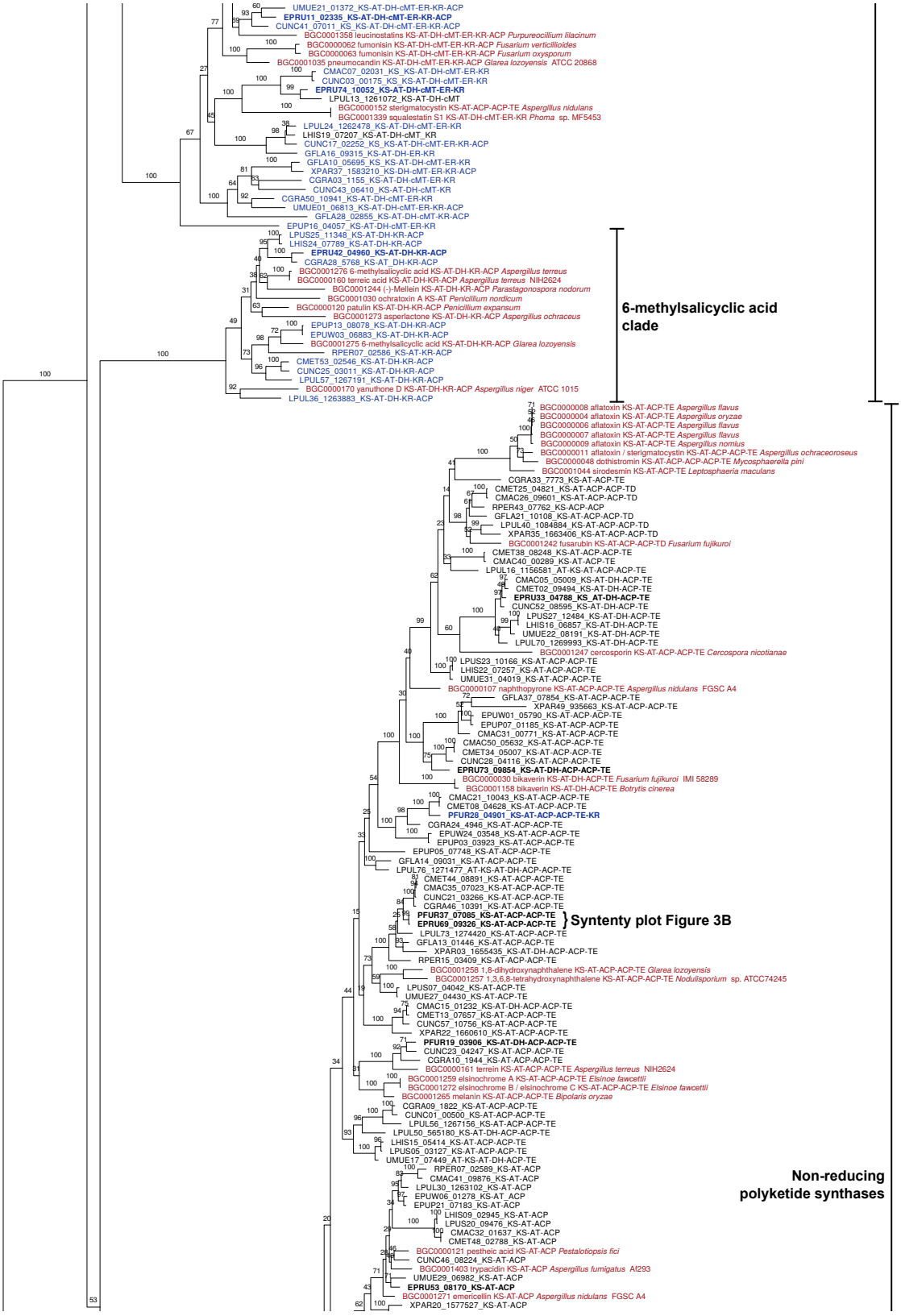
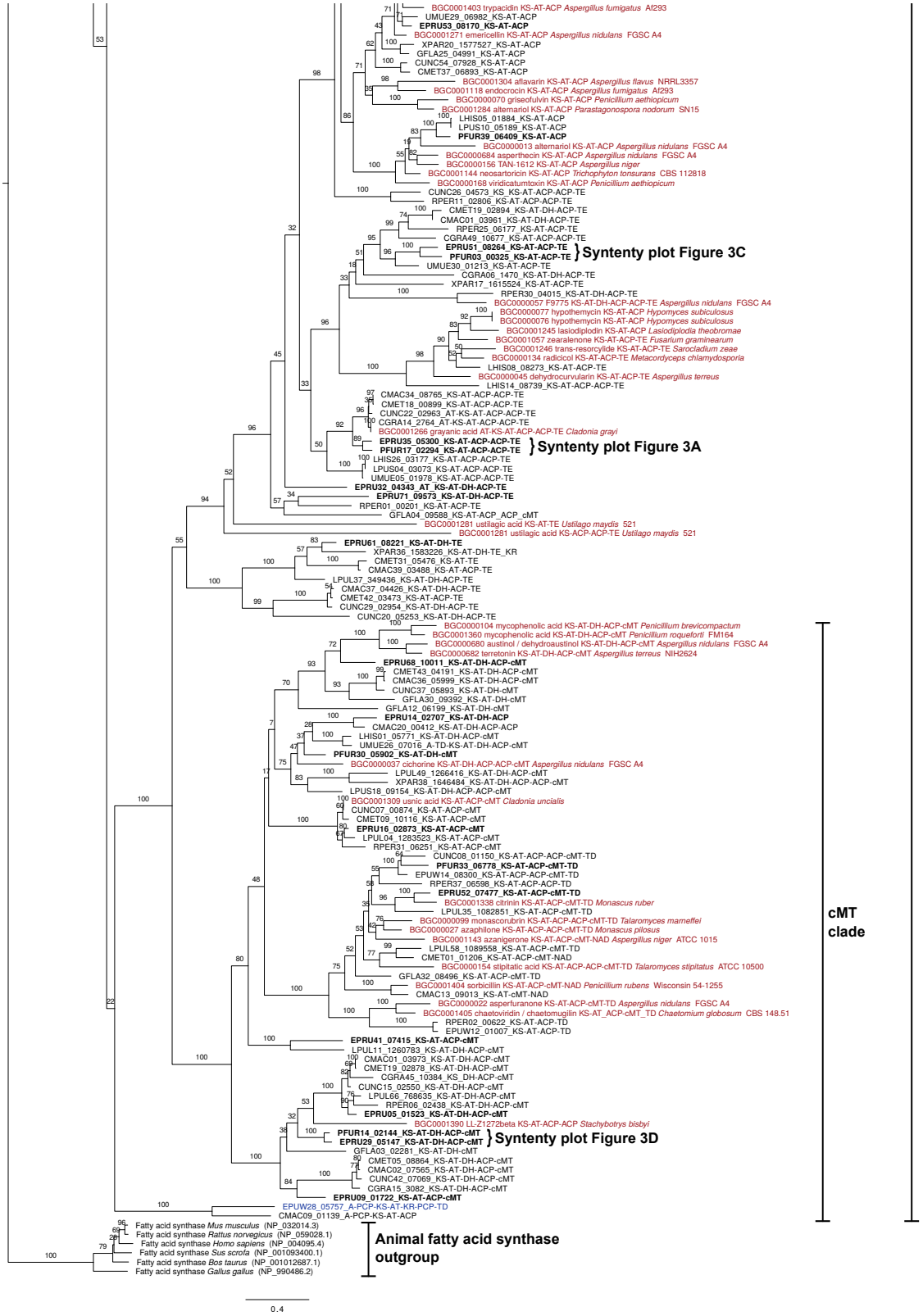


Figure S1. – *continued (Detail page 4 of 4)*



Acknowledgement

I am grateful to all the people that supported me and contributed to this work. First of all, I want to thank my supervisor Prof. Imke Schmitt, for giving me the opportunity to join her working group, for helpful discussions, ideas and guidance. I also kindly thank Prof. Markus Pfenniger for agreeing to review this work. My special thanks go to Francesco Dal Grande for his supervision, permanent support, patience and guidance in difficult times, and motivating me during my PhD. I thank Jürgen for all the lab work he contributed to this work and nice talks. I also want to thank Ricarda Prinz warmly for all the coordinating work, but most of all for her friendship and constant support. I thank all other current and former members of the working group of Imke Schmitt for creating a nice atmosphere for discussions and lunch breaks – it is and was pleasant working with all of you. I also thank Claus Weiland for his computer cluster support. I am also thankful to my friends and family for supporting me, for their interest in my work and patience. Last but not least, I am deeply grateful to Iannis and Levi for their patience and love - you mean the world to me.

Zusammenfassung

Pilze produzieren eine große Anzahl von Naturstoffen, von denen viele biologische Aktivitäten aufzeigen. Unter den Naturstoffen von Pilzen finden sich industriell und medizinisch relevante Substanzen wie zum Beispiel Antibiotika, Immunsuppressiva oder cholesterinsenkende Mittel. Die meisten Leitstrukturen bei der Arzneimittelentwicklung sind von natürlichen Quellen inspiriert oder abgeleitet. Durch steigende Resistenzen gegen geläufige Antibiotika und die Entwicklung multiresistenter Keime steigt der Bedarf an neuen chemischen Verbindungen, die sich als effiziente Arzneimittel erweisen können. Insbesondere natürliche Quellen sind hier vielversprechend, da ein Großteil der organismischen Biodiversität im Hinblick auf potentielle Arzneimittelstrukturen noch unerforscht ist. Vor allem die Untersuchung weniger gut zugänglicher Organismen aus nicht-traditionellen Quellen hat großes Potential für die Entdeckung neuer Naturstoffe. Die Naturstoffe symbiotischer Organismen zum Beispiel wurden bei der Suche nach neuen Arzneimittelleitstrukturen bislang vernachlässigt. Innerhalb der Pilze zeigen die lichenisierten Pilze eine besondere Vielfalt an Naturstoffen, zum Beispiel innerhalb der Stoffklasse der Depside und Depsidone.

Flechten werden meist als symbiotische Interaktion zwischen einem Pilzpartner, dem Mykobionten, und einem oder mehreren fotosynthetischen Partnern, den Fotobionten, beschrieben. Neuere Mikrobiom-Studien haben jedoch gezeigt, dass Flechten eher wesentlich komplexere Vereinigungen multipler Arten darstellen und zusätzlich vielfältige eukaryotische und prokaryotische Organismen beherbergen. Flechtensymbiosen sind in einer Vielzahl von teils extremen Lebensräumen weltweit zu finden, was unter anderem durch die Produktion ihrer einzigartigen und diversen Flechtenstoffe ermöglicht wird. Viele der mehr als tausend charakteristischen Sekundärmetabolite werden vom Mykobionten synthetisiert und weisen vielfältige biologische Aktivitäten auf, die von pharmakologischem Interesse sind. Allerdings erschweren vor allem zwei Aspekte die Nutzbarmachung von Flechtenstoffen im größeren Stil: Flechten weisen eine langsame Wachstumsrate auf und die Kultivierung individueller Partner ist oft nicht möglich oder praktikabel. Daher sind experimentelle Ansätze zur Charakterisierung der

molekularen Mechanismen hinter der Biosynthese vielversprechender Flechtenstoffe stark limitiert.

Abhilfe schafft hier ein genomischer Ansatz, bei dem mit bioinformatischen Methoden Genome auf ihr Biosynthesepotential hin durchsucht werden, um dieses dann anschließend näher zu charakterisieren. Üblicherweise stammen genetischen Information aus der Sequenzierung von Individuen. In flechtenbildenden Pilzen als Teil einer obligaten Symbiose sind aufgrund der Hürden durch langsame Wachstumsraten und Kultivierungsschwierigkeiten andere Ansätze erforderlich. In der Tat stammen die wenigen seit 2011 verfügbaren Genome flechtenbildender Pilze vor allem von Mykobiontenkulturen. Durch die Entwicklung neuer Sequenziertechniken wurde jedoch eine metagenomische Herangehensweise ermöglicht, bei der die genetische Information aus einem Gemisch verschiedener Individuen rekonstruiert wird. Dabei kann die DNA des Zielorganismus vor der Sequenzierung angereichert werden oder direkt sequenziert und anschließend bioinformatisch dem Zielorganismus zugeordnet werden. Die zuletzt genannten Möglichkeiten zur Genomrekonstruktion umgehen die experimentellen Schwierigkeiten im Labor, aber bringt eigene methodische Herausforderungen mit. Da insbesondere in Flechten die genetische Information der Symbionten nicht zu gleichen Anteilen vorliegt, ergeben sich Assemblierungsschwierigkeiten durch die ungleiche Sequenziertiefe und Abdeckung der Genome. Zudem sind spezielle Methoden erforderlich, die Sequenzen nach der Assemblierung taxonomischen Gruppen zuordnen. Für Flechten steckt diese metagenomische Strategie noch in den Anfängen und bisher sind nur wenige Studien veröffentlicht, die metagenomische und auch metatranskriptomische Daten von Flechten nutzen um langwierige Kultivierungsversuche zu umgehen.

Obgleich Flechtenstoffe chemisch sehr divers sind, entstammen die meisten Metabolite einigen wenigen Biosynthesewegen. Insbesondere hervorzuheben sind hier vor allem Polyketidesynthasen, die in der Biosynthese der meisten bekannten Flechtenstoffe involviert sind. Pilzliche Polyketidesynthasen bauen aus einfachen Bausteinen komplexe Moleküle auf, indem sie in mehreren Reaktionszyklen den gleichen Satz katalytischer Enzymdomänen für die Molekülverlängerung und chemische Modifikation verwenden. Durch die iterative Verwendung der Enzymdomänen in pilzlichen Polyketidesynthasen ist es bisher nicht möglich, aufgrund der im Polyketidesynthase-Gens codierten Domänenarchitektur die Struktur des synthetisierten Metabolites her vorzusagen. Infolgedessen ist die funktionelle Charakterisierung von Polyketidesynthasen in Pilzen abhängig von experimentellen Beweisen durch zum Beispiel heterologe Expressions- oder Gen-Knockout-Versuchen. Da flechtenbildende Pilze für experimentelle Gen-

manipulation nicht zugänglich sind, kann die Tatsache genutzt werden, dass Gene, die spezifische Sekundärmetabolite codieren, in der Regel auf zusammenhängenden Genclustern angeordnet sind. Solche Biosynthese-Gencluster können mit Hilfe von Algorithmen im Genom identifiziert werden und haben zur Entdeckung unzähliger Cluster geführt. Bei den meisten dieser Gencluster handelt es sich um sogenannte kryptische Cluster, da weder ihre Funktion noch die codierten Metabolite bisher bekannt sind. Neben der beschriebenen verborgen intrinsischen Programmierung in pilzlichen Polyketidesynthasen selbst, fehlt oft das Hintergrundwissen zu Selektivität und Aktivität von begleitenden Clustergenen oder Genclusterregulatoren und erschwert so die Charakterisierung. Bisher wurde nur ein Bruchteil der entdeckten Biosynthese-Gencluster experimentell evaluiert. In flechtenbildenden Pilzen ermöglichen die oben beschriebenen experimentellen Hürden bisher nur putative Assoziationen von Metaboliten und codierenden Genen.

Zielsetzung dieser Arbeit ist es, einen metagenomischen Ansatz für die Genomrekonstruktion lichenisierter Pilze zu implementieren, um das biosynthetische Potential in dieser Organismengruppe untersuchen zu können. Dabei sollen die Grundlagen geschaffen werden, einen Link zwischen Sekundärmetaboliten und Genen zu ermöglichen und das biosynthetische Potential von Flechten für die Biotechnologie zugänglich zu machen.

Im ersten Teil dieser Arbeit wurden die aus metagenomischen Daten abgeleiteten Pilzgenome und -gene mit Referenzgenomen und -gene aus Sequenzdaten von Mykobiontenkulturen der gleichen Art verglichen. Zudem wurde evaluiert wie sehr das Vorhandensein von Fremd-DNA, die nicht vom Zielorganismus des Mykobionten stammt, unsere Fähigkeit beeinflusst das Mykobiontengenom zu rekonstruieren. Im zweiten Teil dieser Arbeit geht es um die Untersuchung des genetischen Biosynthesepotentials in zwei vielversprechenden Systemen. Bei dem ersten System handelt es sich um nahe verwandte Schwesterarten, die sich in ihrer Ökologie, Verbreitung und Fortpflanzungsstrategie unterscheiden, jedoch ein gleiches chemisches Profil aufweisen. Im zweitem System werden zwei Arten untersucht, die besonders reich an Metaboliten sind und zu den wenigen Arten mit ökonomischer Bedeutung gehören.

Für die Evaluierung eines metagenomischen Ansatzes im Vergleich zur direkten Sequenzierung von Mykobionten, wurden zwei Arten (*Evernia prunastri* und *Pseudevernia furfuracea*) ausgewählt, von denen eine Mykobiontenkultur zur Verfügung stand und die reich an Metaboliten sind. Zunächst wurden Referenzgenome von Kultursequenzen assembliert und mit Hilfe von metagenomischen Transkriptomdaten Gene annotiert.

Die Genomassemblierung und taxonomische Zuordnung der resultierenden Sequenzen wurde an sechs verschiedenen Genomassemblern mit jeweils zwei verschiedenen Methoden zur taxonomischen Gruppierung der Sequenzen getestet. Verglichen wurden die Ergebnisse im Hinblick auf verschiedene Genomstatistiken, Anzahl der Assemblierungsfehler und die Übereinstimmung mit dem Referenzgenomen und -genen. Es ergaben sich große Qualitätsunterschiede zwischen den Genomassemblern, die vermutlich auf stark unterschiedliche Abdeckung und Abundanzen der Sequenzen im Datensatz zurückzuführen sind. Die taxonomische Klassifizierung von Sequenzdaten zu Pilzen wurde mit zwei verschiedenen Methoden beurteilt. Ein Ansatz (MEGAN) platziert mit Hilfe von Algorithmen Sequenzen aufgrund ihrer Ähnlichkeit auf der niedrigst möglichen phylogenetischen Stufe (lowest common ancestor). Der andere Ansatz (MetaWatt) macht sich Genomsignaturen wie Tetranukleotidfrequenzen und die Genomabdeckung zunutze. Beide Methoden ergaben keine wesentlichen Unterschiede in der Qualität der rekonstruierten Pilzengenome. Die Klassifizierung anhand von Sequenzähnlichkeit war jedoch etwas präziser wenn auch rechnerisch aufwendiger. Die Vollständigkeit der Pilzgenome und Pilzgene wurde mit Hilfe eines Datensatzes von Abstammungs-spezifischen orthologen Gene abgeschätzt. Dabei zeigte sich, dass die metagenomisch rekonstruierten Genome und Gene eine mit den Referenzen vergleichbare Vollständigkeit aufweisen. Insgesamt lässt sich erkennen, dass die Assemblierungen aus metagenomischen Daten ähnliche Genomgrößen aufweisen wie die Referenzgenome. Um den Einfluss von Fremd-DNA zu beurteilen, wurde eine manuell kurierte Datenbank mit unseren Referenzgenomen verwendet, um die sequenzierten kurzen DNA-Abschnitte taxonomisch zu klassifizieren. Dreiviertel der DNA-Abschnitte fallen dabei auf den Mykobionten, während andere Pilze mit weniger als ein Prozent vertreten sind. Im nächsten Schritt wurde der Vergleich auf die Gene erweitert und zum einem die Orthologie der Gene miteinander verglichen und zum anderen Biosynthesegene identifiziert und als wichtige Genfamilienklasse der Flechten verglichen. Die Gene der metagenomisch-rekonstruierten Assemblierung überlappen zu 88-90 % mit ihren Referenzgenen. Auch die meisten Biosynthesegene, die in den Referenzgenomen identifiziert wurden, konnten auch aus den metagenomischen Daten rekonstruiert werden. Jedoch gab es auch einige Sekundärmetabolit-Gene, die ausschließlich in den metagenomischen Daten gefunden wurden. Diese könnten auch im Referenzgenom zwar vorhanden aber nicht identifiziert worden sein oder eventuell von anderen Pilzen abstammen. Tatsächlich waren in einer Assemblierung aller Pilzsequenzen, die nicht vom Mykobionten stammen, ein Biosynthese-Gencluster in jeder untersuchten Art zu finden. Die taxonomische Einordnung der kurzen DNA-Sequenzen hat jedoch gezeigt,

dass dieser Anteil von pilzlicher Fremd-DNA nur gering ist. Abschließend lässt sich sagen, dass der hier verwendete metagenomische Ansatz eine zuverlässige Alternative ist, um das Genom flechtenbildender Pilze aus der direkten Sequenzierung von Flechtenthalli zu rekonstruieren. Diese Methode ist insbesondere vielversprechend für nicht kultivierbare eukaryotische Organismen.

Der zweite Teil der Arbeit beschäftigt sich mit der Untersuchung des genetischen Biosynthesepotentials in zwei vielversprechenden Systemen. Dabei sollen alternative Grundlagen geschaffen werden, einem Link zwischen Sekundärmetaboliten und Genen näher zu kommen, da Flechten experimentell nur wenig zugänglich sind. Im ersten System wurden die Biosynthese-Gencluster in den Schwesterarten *Umbilicaria hispanica* und *U. pustulata* untersucht. Dafür wurde ein Referenzgenom aus einer Mykobiontenkultur von *U. hispanica* assembliert und mit dem vorhandenen Genom von *U. pustulata* auf Syntenie verglichen. Die Gene wurden auf Orthologie zwischen den Arten und innerhalb der Biosynthesegene untersucht. Obgleich das Genom von *U. hispanica* weniger kontinuierlich assembliert werden konnte, zeigte sich ein hoher Grad an Syntenie im Genom und in der Anordnung der Gene. Wir haben 18 Biosynthese-Gencluster in *U. hispanica* und 21 in *U. pustulata* identifiziert – bei weitem mehr als aufgrund des chemischen Profils angenommen werden kann. Elf dieser Cluster weisen eine hohe Homologie zwischen den Cluster beider Arten auf. Diese stellen vielversprechende Kandidaten für die Biosynthese gemeinsamer Sekundärmetabolite dar. Dahingegen sind die Cluster, die sich zwischen den beiden Arten unterscheiden, interessant um Einblicke in die Biosynthese von Metaboliten zu ermöglichen, die eventuell eine Rolle in der unterschiedlichen Ökologie und Fortpflanzungsstrategie beider Arten spielen.

Im zweiten System für die Untersuchung des Biosynthese-Potentials in Flechten, verwenden wir die Genome von *E. prunastri* und *P. furfuracea* aus dem ersten Teil dieser Arbeit. Es wurden alle Biosynthese-Gencluster in beiden Arten und zusätzlich in 13 weiteren Genomen von flechtenbildenden Pilzen identifiziert und das Biosynthesepotential verglichen. Zudem wurden die Anzahl der Biosynthesegene in Flechten mit 57 Pilzgenomen aus allen wichtigen Pilzklassen verglichen. Alle Genome der flechtenbildenden Pilze zeigten eine hohe Anzahl an Biosynthese-Genclustern im Vergleich zu nicht-lichenizierten Pilzen. In *E. prunastri* wurden 80 Cluster identifiziert und in *P. furfuracea* 51. *Evernia prunstari* zeigte die höchste Anzahl an Polyketidesynthase-Genen von allen untersuchten Genomen. Des weiteren wurde eine umfassende Charakterisierung von orthologen Gencluster aus *E. prunastri* und *P. furfuracea* in einem vergleichenden genomischen Ansatz erstellt. Die Identifizierung von Regulatorgenen in den

Genclustern könnte entscheidend in heterologen Expressionversuchen sein. Homologe Cluster sind gute Kandidaten für Metabolite, die in beiden Arten vorkommen. Es wurde eine Phylogenie erstellt, die auf einer konservierten Domäne der Polyketidesynthasen basiert. Dafür wurden allen untersuchten Flechtengenome mit einem Datensatz von funktionell charakterisierten pilzlichen Polyketidesynthase zusammen gebracht. Mit Hilfe eines solchen Ansatzes lassen sich interessante Cluster identifizieren, die phylogenetische Nähe zu charakterisierten Genen aufweisen. Das erkundete Biosynthese-Potential in allen untersuchten Flechten übersteigt bei weitem die Anzahl, die aufgrund der chemischen Profile erwartet werden kann. Dies hebt das enorme Potential hervor, dass in Flechten schlummert und darauf wartet entdeckt und funktionell charakterisiert zu werden, wobei vergleichende Genomik und phylogenetische Analysen wertvolle Hinweise liefern können. Der implementiere metagenomische Ansatz ist vielversprechend um auch das Potential von nicht kultivierbaren Mykobionten oder anderen schwer zugänglichen Eukaryoten zu erfassen. Diese Arbeit legt wichtige Grundlagen, um das aufgezeigte Potential von Flechten eines Tages nutzbar zu machen.