The Lichenologist **50**(3): 329–340 (2018) © British Lichen Society, 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

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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*;

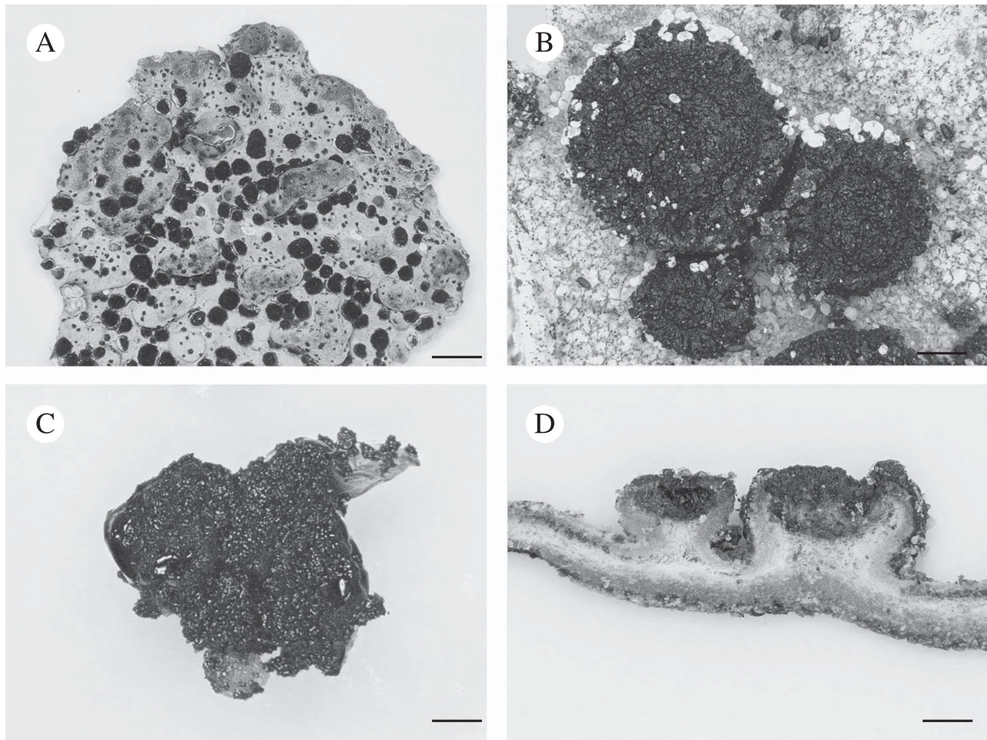

FIG. 1. *Lasallia hispanica*. A, thallus with apothecia; B, apothecia; C, mycobiont culture used for genome sequencing; D, section of thallus with apothecia. Scales: A = 10 mm; B–D = 1 mm.

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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.

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

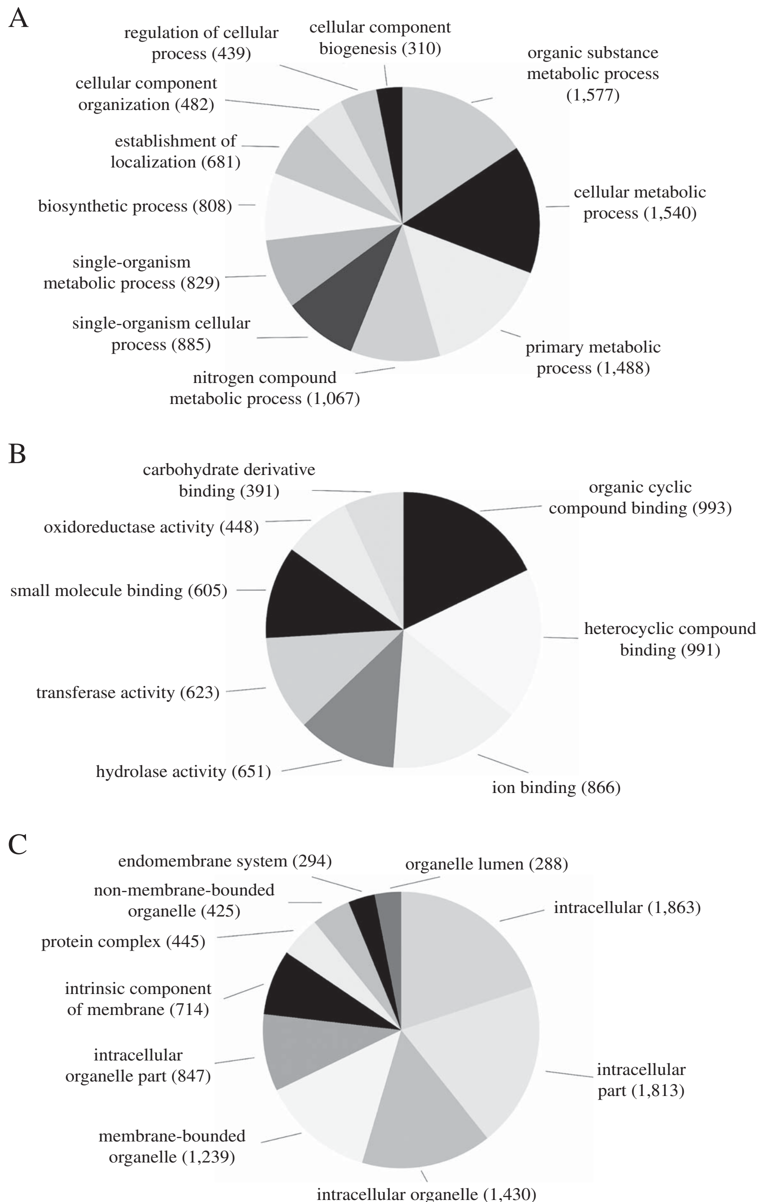

FIG. 2. Distribution of Blast2GO annotations for *L. hispanica*. Charts show level 3 annotations for Biological Process (A), Molecular Function (B) and Cellular Components (C).

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

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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),

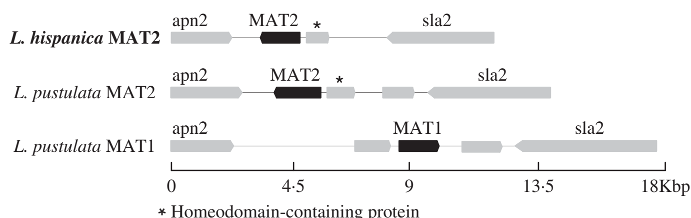

Fig. 3. Configuration of the *MAT* loci in *L. hispanica* and in two *L. pustulata* assemblies (*MAT1*: European Nucleotide Archive GCA_000938525.1; *MAT2*: Greshake (2018)).

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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	–	–

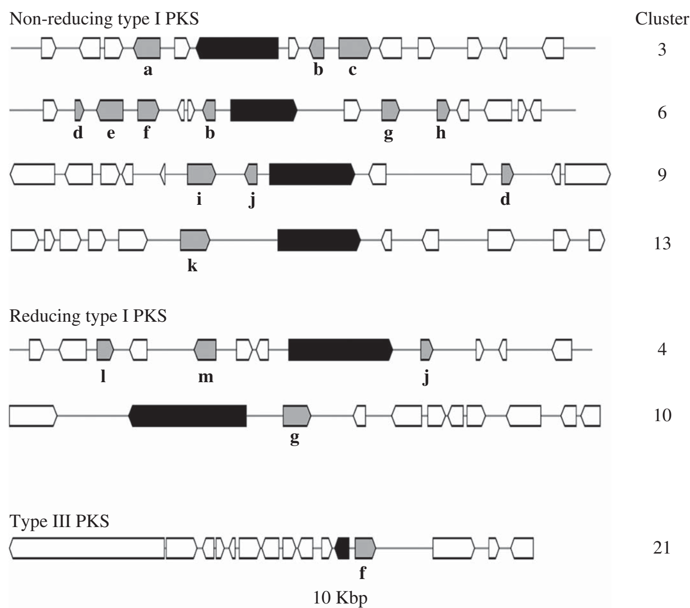

FIG. 4. Configuration of *L. hispanica* biosynthetic gene clusters with orthologs in *L. pustulata*. Black boxes represent core biosynthetic genes (PKSs in the upper six clusters and a chalcone and stilbene synthase in the bottom cluster). Shaded boxes indicate genes coding for tailoring enzymes: a, 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).

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

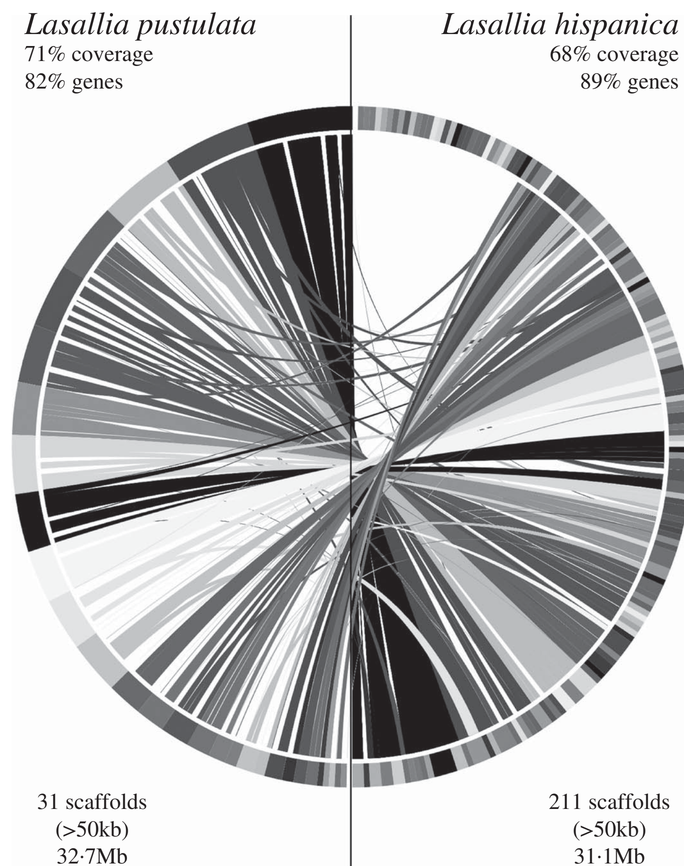

FIG. 5. Circle plot of the genome alignment between 31 *L. pustulata* (left) and 211 *L. hispanica* (right) scaffolds. Scaffolds of *L. hispanica* were ordered to align against the genome of *L. pustulata* using information from 202 syntenic blocks.

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

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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.

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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)

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press

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.

https://doi.org/10.1017/S002428291800021X Published online by Cambridge University Press