**Characterizing the ribosomal tandem repeat in a clade of symbiotic fungi**

Michael Bradshaw1, Felix Grewe2, Anne Thomas1, Cody H. Harrison1, Hanna Lindgren2, Lucia Muggia3, Larry St. Clair1, H. Thorsten Lumbsch2, Steven D. Leavitt1,\*

**Abstract**

Regions within the nuclear ribosomal operon are a major tool for inferring evolutionary relationships and investigating diversity in fungi. In spite of the prevalent use of ribosomal markers in fungal research, central features of ribosomal evolution are poorly characterized for fungi in general, including lichenized fungi. The internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (nrDNA) has been adopted as the primary DNA barcode identification marker for fungi. However, little is known in regard to intragenomic variation in nrDNA. In order to better understand evolution of nrDNA and the utility of the ITS for barcode identification of lichen-forming fungal species, we generated nearly complete nuclear ribosomal operon sequences from species in the *Rhizoplaca melanophthalma* species complex using short reads from high-throughput sequencing. We estimated copy numbers for the nrDNA operon, ranging from nine to 48 copies for members of the *R. melanophthalma* complex, and find low levels of intragenomic variation in the standard barcode region (ITS). The limited number of potentially polymorphic sites largely do not correspond to fixed diagnostic nucleotide position characters separating taxa within this species complex, supporting the role of the ITS as an effective barcoding marker. Monophyly of currently described species in this complex was supported in phylogenetic reconstructions of the ITS, 28S, IGS, and some intronic regions; however, phylogenetic reconstructions based on the 18S provided much lower resolution. Phylogenetic analysis of concatenated ITS and intergenic spacer sequence data generated from 496 specimens collected worldwide revealed, previously unrecognized lineages in the nrDNA phylogeny. Whether these lineages represent undescribed species-level lineages or reflect uncharacterized aspects of nrDNA evolution remains unknown.

**Key words:** copy number variation, DNA barcoding, ITS, lichens, nrDNA, repeat region

**Introduction**

For eukaryotes, regions within nuclear ribosomal (nrDNA) operon have been instrumental in characterizing diversity and inferring evolutionary relationships (White et al., 1990; Wurzbacher et al.). The eukaryotic nuclear ribosomal operon is arranged in tandem repeats in the nuclear genome, with each repeat containing genes, various spacer regions, introns, and other less understood elements (Gutiérrez et al., 2007; Hillis and Dixon, 1991). Copy number of the nrDNA operon is a rapidly evolving trait (Szostak and Wu, 1980). Across Fungi, nrDNA copy number has been shown varied considerably, ranging from tens to over 1400 copies per genome (Lofgren et al.). The relative ease of amplification, coupled variable substitution rates among different regions of nrDNA, have promoted its longstanding use in phylogenetic and biodiversity research (Wurzbacher et al.).

The internal transcribed spacer region (ITS: comprising ITS1, 5.8, and ITS2) within the nrDNA operon has been proposed as the standard DNA barcoding region for Fungi (Schoch et al., 2012). The ITS barcode currently plays a fundamental role in characterizing fungal diversity (Nilsson et al., 2019). However, the integrity of the ITS region for barcoding is questionable due to studies reporting conflicting results as to the levels of intragenomic variation found within the fungal nrDNA operon (Lindner and Banik, 2011). Some studies report that mutations and variation within the ITS region are relatively minor and little practical consequence, attributing the consistency to concerted evolution (Ganley and Kobayashi, 2007). Other studies, however, report significant amounts of variation, suggesting that evolution may not occur in a purely concerted manner (Simon and Weiß, 2008). Studies using traditional Sanger sequencing may effectively conceal potential intragenomic variation due to PCR bias or dominating signal from the predominantly amplified copy. Cloned sequencing studies have revealed the occurrence of intragenomic variation in nrDNA in multiple fungal lineages (Harrington et al., 2014; Lindner and Banik, 2011). In fact, a pyrosequencing-based study suggests that distinct ITS copies are found in multiple Ascommycete and Basidiomycete lineages, although in relatively low proportions (Lindner et al., 2013).

Impetus for including the full nrDNA operon in published genome assemblies (Lindner and Banik, 2011) and use of nrDNA in emerging long‐read sequencing technologies (Wurzbacher et al.) highlights the need of improved characterization of this important genomic region. While information from nrDNA has been fundamental in research into lichen-forming fungi (DePriest, 1993; Gargas et al., 1995; Gutiérrez et al., 2007; Myllys et al., 1999; Thell, 1999), including DNA barcoding studies (Divakar et al., 2016; Kelly et al., 2011; Leavitt et al., 2014), the full nrDNA operon has not yet been characterized for any members of these important symbiotic fungi that we are aware of. Furthermore, in spite of suggestive evidence that some lineages of lichenized fungi harbor multiple distinct copies of the ITS region (CITATIONS), intragenomic nrDNA variation has not yet been explicitly tested.

In this study we investigated members of the *Rhizoplaca melanophthalma* species complex (Leavitt et al., 2011) to better understand features of nrDNA in lichen-forming Ascomycetes. The *Rhizoplaca melanophthalma* species complex monophyletic lineage comprised of ca. ten closely-related species/species-level lineages that originated during the Miocene and diversified largely during and Pleistocene (Leavitt et al., 2013b). Previous empirical species delimitation studies have circumscribed robust species boundaries among closely-related and morphologically similar species (Leavitt et al., 2011; Leavitt et al., 2013a) which have largely been supported by genome-scale molecular data (Grewe et al., 2017; Leavitt et al., 2016). All formally described species in the complex can be identified using the standard DNA barcoding marker, the ITS (Leavitt et al., 2013a), with the exception of the two vagrant species *R. haydenii* (Tuck.) W. A. Weberand *R. idahoensis* Rosentreter & McCune.

To more fully characterize the nuclear ribosomal operon in the *Rhizoplaca melanophthalma* species complex, we (i) generated nearly complete assemblies of the nuclear ribosomal operon from short-read, high-throughput sequencing data, (ii) estimated the number of copies of the ribosomal operon repeat region, (iii) assessed the range of intragenomic variation in the ITS region – the formal DNA barcoding marker in fungi – using high-throughput sequencing data, and (iv) compared topologies inferred from different regions of the nuclear ribosomal operon. The results of this study provide valuable insight into the utility of this rDNA as a barcoding marker for fungi.

**2. Materials and Methods**

*2.1. Taxonomic sampling and data compilation*

Our sampling included representatives of the nine formally recognized species within the *R. melanophthalma* species complex (Leavitt et al., 2013a) and two outgroup taxa – *R. subdiscrepans* (Nyl.) R. Sant*.* and *Protoparmeliopsis peltata* (Ramond) Arup, Zhao Xin & Lumbsch (supplementary Table S1). For this study, we analyzed short-read metagenomic data from a total of 33 specimens, representing ten *Rhizoplaca* s. lat. species (Leavitt et al., 2016a), including: *R. haydenii* (Tuck.) W.A. Weber (n=2), *R. melanophthalma* (DC.) Leuckert 1977 (n=7), *R. novomexicana (H. Magn.)* S.D. Leav., Zhao Xin & Lumbsch (n=1), *R. parilis* S.D. Leav. (n=4), Fern.-Mend., Lumbsch, Sohrabi & St. Clair, *R. occulta* S.D. Leav. (n=2), Fern.-Mend., Lumbsch, Sohrabi & St. Clair, *R. polymorpha* S.D. Leav., Fern.-Mend., Lumbsch, Sohrabi & St. Clair (n=6), *R. porteri* S.D. Leav., Fern.-Mend., Lumbsch, Sohrabi & St. Clair (n=5), *R. shushanii* S.D. Leav., Fern.-Mend., Lumbsch, Sohrabi & St. Clair (n=5), and single representative of two outgroup taxa – *R. subdiscrepans* and *P. peltata*.

In order to more fully characterize the range of ITS diversity in the *Rhizoplaca melanophthalma* species complex, amplicon-based sequence data was generated from an additional XX specimens from the *R. melanophthalma* species complex collected throughout western North America, the center of diversity for this group (Leavitt et al., 2013b). For all new specimens, DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega), and amplification and sequencing of the ITS marker followed previously described methods (Leavitt et al., 2011). Newly generated sequences were combined with previously available nrDNA sequence data from the *Rhizoplaca melanophthalma* group (<https://treebase.org/>; study No. 19048), for a total of 496 specimens.

*2.2. Short-read data, genome assembly and identification of the nuclear ribosomal operon*

Short reads from 33 *Rhizoplaca* specimens reported in a previous study (Leavitt et al., 2016a) were used for genome assembly to identify contigs containing the nuclear ribosomal operon (nrDNA). Full details of specimen preparation and sequences are described in (Leavitt et al., 2016a). In short, libraries were prepared using the Illumina Nextera XT DNA library prep kit (product discontinued) and were pooled and sequenced on a single lane of Illumina’s HiSeq2000 platform, generating 100-bp paired-end reads with a 350-bp insert size, with the exception of four specimens that were sequenced individually on the MiSeq platform (Illumina) generating 250-bp paired-end (PE) reads with a 550-bp insert size. All paired-end (PE) reads were filtered using TRIMMOMATIC v0.33 (Bolger et al., 2014) before assembly to remove low quality reads and/or included contamination from Illumina adaptors using the following parameters: ILLUMINACLIP; LEADING:3; TRAILING:3; SLIDINGWINDOW:4:15; and MINLEN:36. *De novo* genome assemblies were constructed using SPAdes v3.5.0 (Nurk et al., 2013) running a single read error correction iteration prior to the genome assembly using kmer values of 55, 77, 99, with the mismatch careful mode (--careful) enabled, and using MismatchCorrector. From each assembly, contigs containing the nrDNA were identified using a custom BLAST (Altschul et al., 1990) search implemented in the program Geneious R11 (Kearse et al., 2012) against available regions of the nrDNA, e.g., nuLSU, IGS, and ITS, generated from *Rhizoplaca melanophthlama s. lat.* specimens.

While reads from the axenic reference culture (‘mela\_REF’) were exclusively derived from the targeted *R. melanophthalma* fungal genome, genomic libraries prepared from all field-collected specimens were comprised not only of DNA from the targeted mycobiont, but also DNA from the complete holobiont, e.g., associated *Trebouxia* photobiont, secondary fungi, bacteria, etc. (Arnold et al., 2009; Cardinale et al., 2008; Hodkinson and Lutzoni, 2009; Muggia et al., 2013). Some regions of nrDNA are highly conserved across divergent lineages (e.g. 18S, 5.8S and portions of 28S subunits), and reads from non-target genomes may potentially bias the interpretation of intragenomic variation within *Rhizoplaca* species. Therefore, we used a *de novo* assembly approach of all nrDNA reads to separate nrDNA cluster of the targeted *Rhizoplaca* mycobiont from reads of other symbionts that co-occur within lichens. For each specimen, PE reads were mapped back to its respective contigs containing nrDNA from the SPAdes assembly using the Geneious R11 Read Mapper, with the “medium-low sensitivity/ fast” settings, iterated 5 times. Successfully mapped reads were then assembled *de novo* using the native Geneious R11 Assembler at ‘medium-low sensitivity’. Resulting contigs were searched against NCBI’s GenBank database using BLAST to identify non-target contigs; and the non-target contigs were excluded from further analysis.

*2.3. Assessing intragenomic variation of the ITS, inferring copy number of ribosomal operon, and intron identification*

Our assessment of potential intragenomic variation focused on this the ITS region (ITS1, 5.8S, and ITS2) – the standard DNA barcoding marker for specimen identification for fungi (Schoch et al., 2012a). To identify potentially polymorphic sites in the nrDNA, PE reads from each specimen were mapped back to their corresponding ITS region extracted from the Geneious assembly, with a 600 bp buffer on either end, using the BWA (Li and Durbin, 2009). The 600 bp buffer on either end was used to ensure that all reads containing portions of the ITS region were indeed mapped back to the reference rather than being discarded because part of the read mapped to a region of the ribosomal operon before or after the ITS region. Samtools v1.6 genomics utilities package (Li et al., 2009) was used to process alignment output, filtering out unmapped reads so that only reads corresponding to the ITS and bordering regions remained. A samtools pileup file was generated to identify the bases aligned with each position of the reference sequence, visually confirmed by the Geneious v11 and Integrative Genome Viewer (Robinson et al., 2011). A python script identified mismatches and calculated percent variance at each position in the pileup file. To calculate percent variance, the number of reads that varied from the consensus at that location was divided by coverage at that location. When calculating the percent variance, there was no effort taken to identify bases within a read that were sequencing error or true variation. The idea being that sequencing error and true variation would be distinguishable based on the percent variance. Sequencing error would be comparable to known error rates of the sequencing technology (1% for Illumina) while intragenomic variation would result in percent variance much higher.

To estimate the total copy number of the ribosomal operon, we compared average read depth coverage of nrDNA relative to coverage of putative single-copy regions of the nuclear genome. For each specimen, reads were mapped back to its respective Geneious assembly of the nrDNA operon, the three largest contigs (ca. 363 kb, 307kb, and 227 kb, respectively) from the draft genome assembly from the axenic culture (Leavitt et al., 2016a), and known single-copy genes, *MCM7*, *RPB1* and *RPB2*. The average coverage depth of the nrDNA operon was divided by the average coverage depth of the nuclear single copy genes and nuclear genomic regions. The difference in coverage was interpreted as an approximation for the copy number of the nuclear ribosomal operon.

In contrast to most other eukaryotic genomes, yeast genomes have few introns (Spingola et al., 1999). Therefore, we used a nrDNA sequence from *Saccharomyces paradoxus* (GenBank accession No. BR000309) to identify introns and demark boundaries between the 18S, ITS1, 5.8S, ITS, 28S, and IGS regions. No attempt was made to distinguishing different intron types – e.g., group I, group II, and spliceosomal introns.

A group I intron at the 3’ end of the SSU has previously been shown to present in all species within the *R. melanophthalma* group, except *R. porteri* (Leavitt et al., 2011); and the absence of this intron served as a diagnostic character in the description of this taxon (Leavitt et al., 2013a). However, PCR amplifications may not provide an accurate perspective of repetitive genomic regions due to PCR bias or overwhelming signal from the most commonly amplified variant. Therefore, to verify the absence of this group I intron, we attempted to map reads from *R. porteri* specimens to a consensus sequence representing this intron using the Geneious v11 Read Mapper, with the “medium-low sensitivity/ fast” settings, iterated 5 times. To test if this group I intron may be absent in some copies of rDNA in other species in the *R. melanophthalma* group, we searched PE reads from all *R. haydenii*, *R. melanophthalma*, *R. parilis*, *R. polymorpha*, *R. porteri*, *R. occulta*, and *R. shushanii* specimens for the conserved motif lacking the intronusing a custom script.

*2.4 Multiple sequence alignments and phylogenetic reconstructions*

An initial multiple sequence alignment (MSA) of the nearly complete nrDNA operon assembly (n=33) was performed using the program MAFFT v7 (Katoh et al., 2005; Katoh and Toh, 2008), implementing the FFT-NS-i iterative refinement method. To improve alignment accuracy for specific phylogenetic comparisons of different regions within the ribosomal operon, individual alignments were constructed independently for the 18S, ITS (ITS1, 5.8S, ITS2), 28S, the IGS, and each intron present in the 28S and 18S regions. After excluding introns, MSAs of the 18S, 28S rDNA, and ITS region were aligned in MAFFT using the G-INS-i algorithm The IGS and intronic regions were aligned individually using the E-INS-i algorithm for sequences with conserved domains and long gaps.

Previous studies have indicated that species in the *R. melanophthalma* species complex can be discriminated using phylogenetic reconstructions of the barcoding marker for fungi, the ITS region (Leavitt et al., 2011; Leavitt et al., 2013a); and here we investigated if species within this complex are also recovered as monophyletic using other regions of nrDNA for phylogenetic inference. We reconstructed phylogenies from different regions of the ribosomal operon: (i) 18S rDNA, excluding introns; (ii) introns within the 18S region; (iii) 28S rDNA, excluding introns; (iv) introns within the 28S region; (v) concatenated 18S and 28S rDNA, excluding introns; (vi) concatenated introns from both 28S and 18S regions, (vii) the IGS region; and (viii) a complete matrix comprised of 18S and 28S rDNA, and associated introns, and the IGS region. Only introns that were present in all of the ingroup samples – the *R. melanophthalma* group – were included in phylogenetic analyses to minimize bias from highly mobile introns that may have been incorporated or lost more recently than the most recent common ancestor of the *R. melanophthalma* group.

Maximum likelihood (ML) topologies were inferred from each of these regions individually using the program RAxML v8.2.2 (Stamatakis, 2006; Stamatakis et al., 2008). All ML analyses were performed using the CIPRES Science Gateway server (http://www.phylo.org/portal2/), using the ‘GTRGAMMA’ model and evaluating nodal support using 1000 bootstrap pseudo-replicates. A ML topology was also inferred from the complete rDNA matrix using RAxML, treating each region (IGS, SSU, ITS, and LSU) as separate partitions; otherwise analyses were performed as described above.

To compare the nrDNA sequence variation of the 33 specimens sampled from the *R. melanophthalma* group within the context of a broader sampling of specimens, we compiled a nrDNA matrix comprised of IGS and ITS sequences from a previous study (Leavitt et al., 2016b) and newly sampled specimens, resulting in a total of 496 specimens (supplementary Table S1). For comparison, a number of specimens were represented by multiple ITS sequences, including those assembled from PE reads for this study and sequences generated using Sanger sequencing from the initial DNA extractions used for high-throughput sequencing library preparation. A ML topology was inferred from this dataset using IQTree v1.6.3 (Nguyen et al., 2015), treating each region (IGS and ITS) as separate partitions. IQTree ML analyses were run under JTT + Empirically counted frequencies from alignment + FreeRate model with 7 categories, the best model identified by ModelFinder, and the commonly used models GTR + Optimized base frequencies + Free rate model with 6 categories and WAG + Optimized base frequencies + Free rate model with 6 categories.

*2.5 Reproducibility*

Sequences and alignments generated and used in this study are available were submitted to TreeBase (<https://treebase.org/>; study No. XXXXX). All custom scripts were deposited at XXX.

***3. Results***

*3.1. Operon assemblies and coverage*

In the initial SPAdes assemblies, the nrDNA operon was assembled into a single contig comprised of the complete 18S, ITS1, 5.8S, ITS2, and 28S regions, and most of the IGS region. However, in all specimens, a short region near the 5’ end of the intergenic spacer region (IGS) remained ambiguous due to a region of repeats expanding beyond the libraries’ insert size. Of the reads mapped to the initial nrDNA SPAdes assemblies, between ca. 85–92% were assembled as one contig representing the mycobiont nrDNA operon, 4–6% as contigs representing the photobiont nrDNA operon, and remaining reads, ca. 3–7%, were assembled into small, low coverage contigs.Subsequent *de novo* Geneious assemblies using only reads initially mapped back to the original nrDNA contig from the SPAdes assembly were highly congruent with the original assemblies. For libraries run on the HiSeq platform, average coverage of the nrDNA operon was 317x, ranging from ca. 79x (‘mela\_8800’) to ca. 950x (poly\_8668g). For the four specimens sequenced on the MiSeq platform, the average coverage of nrDNA was 1324x, ranging from ca. 376x (‘novo\_8664d’) to ca. 1813x (‘subd\_9052’) (Table S1).

In the *R. melanophthalma* group, the length of the nrDNA operon ranged from ca. 11.1–11.2 kb in *R. melanophthalma s. str.* to ca. 14.0–15.9 kb in *R. shushanii* (Table 1). In the outgroup taxa, *P. peltata* and *R. subdiscrepans*,nrDNA operons were considerable shorter – 8.5 kb and 8.9 kb, respectively. Differences in the lengths of the nrDNA operon were largely due to variable intron patterns in the 18S, 28S, and IGS. The aligned nrDNA operon (18S, ITS1, 5.8 S, ITS2, 28S, and IGS) included a total of 18,437 nucleotide position characters.

*3.2. nrDNA operon copy number estimates, intragenomic variation, and introns*

We found low levels of intragenomic variation in the standard DNA barcoding region for fungi – the ITS – in members of the *R. melanophthalma* group, with variance at a single nucleotide position character rarely exceeding 10% (Fig. 1; Supplementary Table S2). Furthermore, potentially polymorphic sites generally did not coincide with segregating sites that separated species (Fig. 1). The estimated copy number of the ribosomal operon ranged from 8.7 (SD=4.6) in *R. shushanii* (‘shus\_8664-3’) to 42.7 (SD=4.1) in *R. porteri* (‘port\_8796’) (Table 1); and estimates were similar across a range of comparisons (Supplementary Table S2). Intraspecific nrDNA copy number variation was observed in all taxa represented by multiple individuals. Furthermore, conspecific specimens collected at close spatial scales (< 100 m.) also showed variation in rDNA copy number. A total of 13 introns were identified within the 18S and ten in the 28S (Supplementary Table SSU & Supplementary Table LSU; TreeBase alignment).

*3.3. Phylogenetic relationships inferred from nrDNA*

Topologies inferred from the ITS (ITS1, 5.8S, and ITS2), IGS, and 28S datasets independently each recovered all species as monophyletic and varying levels of support among species-level lineages (Fig. 2A–C), while the topology inferred from the 18S was poorly resolved (Fig. 2D). Phylogenies inferred from the concatenated intronic regions in the 18S and 28S also recovered all species as monophyletic and generally with strong nodal support (Fig. 2D & F). Relationships among species-level clades varied widely depending on the nrDNA dataset. The ML analyses of the complete rDNA dataset provided a fully resolved, well-supported topology (Fig. 3).The distribution of introns was mapped onto the complete rDNA topology (Fig. 3).

In the comprehensive ITS/IGS topology (n=496) most previously recognized species-level clades were recovered as well-supported monophyletic clades with a few notable exceptions (Fig. 4; Fig. S5). *Rhizoplaca polymorpha* was recovered as monophyletic with weak statistical support; and *R. haydenii*, which was recovered in two separate clades (Fig. 4). Three previously unrecognized clades were recovered in the ITS/IGS topology – ‘nrDNA clade I’, ‘nrDNA clade II’, and ‘nrDNA clade III’ (Fig. 4).

**Discussion**

The estimated copy number was used as a threshold for determining significant variance. If the percentage of variance is less than the minimum percentage represented by a single copy of the ITS repeat, the variance is considered insignificant. Illumina sequencing typically shows an error rate of approximately 1%,  providing another lower bound on confidence in observed variance.

**Long read technologies**

**Acknowledgements**

**References**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403-410.

Arnold, A.E., Miadlikowska, J., Higgins, K.L., Sarvate, S.D., Gugger, P., Way, A., Hofstetter, V., Kauff, F., Lutzoni, F., 2009. A Phylogenetic Estimation of Trophic Transition Networks for Ascomycetous Fungi: Are Lichens Cradles of Symbiotrophic Fungal Diversification? Systematic Biology 58, 283-297.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Cardinale, M., Jr, J.V.d.C., Müller, H., Berg, G., Grube, M., 2008. *In situ* analysis of the bacterial community associated with the reindeer lichen *Cladonia arbuscula* reveals predominance of *Alphaproteobacteria*. FEMS Microbiology Ecology 66, 63-71.

Ganley, A.R.D., Kobayashi, T., 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Research 17, 184-191.

Grewe, F., Guo, W., Gubbels, E., Hansen, A.K., Mower, J., 2013. Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. BMC Evolutionary Biology 13, 8.

Hodkinson, B., Lutzoni, F., 2009. A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. Symbiosis 49, 163-180.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33, 511-518.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefin Bioinformatics 9, 286-298.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Leavitt, S.D., Fankhauser, J.D., Leavitt, D.H., Porter, L.D., Johnson, L.A., St. Clair, L.L., 2011. Complex patterns of speciation in cosmopolitan ‘‘rock posy’’ lichens – Discovering and delimiting cryptic fungal species in the lichen-forming *Rhizoplaca melanophthalma* species-complex (Lecanoraceae, Ascomycota). Molecular Phylogenetics and Evolution 59, 587-602.

Leavitt, S.D., Fernández-Mendoza, F., Pérez-Ortega, S., Sohrabi, M., Divakar, P.K., Lumbsch, H.T., St. Clair, L.L., 2013a. DNA barcode identification of lichen-forming fungal species in the *Rhizoplaca melanophthalma* species-complex (Lecanorales, Lecanoraceae), including five new species MycoKeys 7, 1–22.

Leavitt, S.D., Fernández-Mendoza, F., Pérez-Ortega, S., Sohrabi, M., Divakar, P.K., Vondrák, J., Thorsten Lumbsch, H., Clair, L.L.S., 2013b. Local representation of global diversity in a cosmopolitan lichen-forming fungal species complex (*Rhizoplaca*, Ascomycota). Journal of Biogeography 40, 1792–1806.

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

Leavitt, S.D., Kraichak, E., Vondrak, J., Nelsen, M.P., Sohrabi, M., Perez-Ortega, S., St Clair, L.L., Lumbsch, H.T., 2016b. Cryptic diversity and symbiont interactions in rock-posy lichens. Molecular Phylogenetics and Evolution 99, 261-274.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Muggia, L., Vancurova, L., Škaloud, P., Peksa, O., Wedin, M., Grube, M., 2013. The symbiotic playground of lichen thalli – a highly flexible photobiont association in rock-inhabiting lichens. FEMS Microbiology Ecology 85, 313-323.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268-274.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel, S., Woyke, T., Tesler, G., Alekseyev, M., Pevzner, P., 2013. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng, M., Jiang, R., Sun, F., Zhang, X. (Eds.), Research in Computational Molecular Biology. Springer Berlin Heidelberg, pp. 158-170.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature Biotechnology 29, 24.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Consortium, F.B., 2012a. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences, doi: 10.1073/pnas.1117018109.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Consortium, F.B., 2012b. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences 109, 6241-6246.

Simon, U.K., Weiß, M., 2008. Intragenomic variation of fungal ribosomal genes is higher than previously thought. Molecular Biology and Evolution 25, 2251-2254.

Spingola, M., Grate, L., Haussler, D., Ares, M., 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA 5, 221-234.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57, 758-771.

**Table 1.** Variation in the size of rDNA operon

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Taxon | Operon length | Estimated copy # | # introns in SSU | # introns in LSU |
| *R. haydenii* (2) | 12.4 kb | 14–21 |  |  |
| *R. melanophthalma* (x) | 11.1–11.2 kb | 12–24 |  |  |
| *R. novomexicana* (1) | 12.4 kb | 16 |  |  |
| *R. occulta* (2) | 12.6–12.7 kb | 27–28 |  |  |
| *R. parilis* (4) | 12.6 kb | 14–36 |  |  |
| *R. polymorpha* (6) | 12.9 kb | 16–28 |  |  |
| *R. porteri* (5) | 11.7–11.8 kb | 19–43 |  |  |
| *R. shushanii* (5) | 14.0–15.9 kb | 9–22 |  |  |
| *P. peltata* (1) | 8.5 kb | x |  |  |
| *R. subdiscrepans* (1) | 8.9 kb | x |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | # of base pairs (bp) | Alignment length | # of putative group I introns | # of putative splice introns |
| IGS | 2041 bp (*R. novomexicana*)– 3087 bp (*R. shushanii*) | 3951 bp | Not evaluated | Not evaluated |
| 18S | 3502 bp (*P. porteri*) – 7082 bp (*R. shushanii*) | 8019 bp | 11 | 1 |
| ITS1 |  |  | NA | NA |
| 5.8S |  |  | NA | NA |
| ITS2 |  |  | NA | NA |
| 28SS | 4628 bp (*R. melanophthalma*) – 5348 bp (*R. polymorpha*) | 5,990 | 7 | 3 |
| rDNA total | 11062–15870 b(ingroup)  8,474–8,912 (outgroup) | bp |  |  |
|  |  |  |  |  |

**Table 2.**

**Table 3.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TAXON/DNA code | Length of operon kb  (IGS/SSU/ITS/LSU) | | Estimated Copy  # (SD) | 95% |  | num |
| R. haydenii/8678 | 4.333333333 | | 21 (4.3) | 2.83 |  |  |
| *R. haydenii*//8935p | 3.833333333 | | 14 (3.8) | 2.50 |  |  |
| *R. melanophthalma* / 8668b | 2.7 | | 17 (2.7) | 1.76 |  |  |
| *R. melanophthalma* / 8800 | 2.7 | | 12 (2.7) | 1.76 |  |  |
| *R. melanophthalma* / 8801 | 3.833333333 | | 24 (3.8) | 2.50 |  |  |
| *R. melanophthalma* / 8802 | 3.166666667 | | 21 (3.2) | 2.07 |  |  |
| *R. melanophthalma* / 8810 | 3.766666667 | | 10 (3.8) | 2.46 |  |  |
| *R. melanophthalma* / REF | 14.26666667 | | 21 (14.3) | 9.32 |  |  |
| *R. novomexicana* / 8684d | 6.7 | | 16 (6.7) | 4.38 |  |  |
| *R. occulta* / 8807\_1 | 4.366666667 | | 28 (4.4) | 2.85 |  |  |
| *R. occulta* / 8807\_4 | 3.666666667 | | 27 (3.7) | 2.40 |  |  |
| *R. parilis* / 8665n | 3.5 | | 20 (3.5) | 2.29 |  |  |
| *R. parilis* / 8803 | 4.266666667 | | 36 (4.3) | 2.79 |  |  |
| *R. parilis* / 8805 | 3.433333333 | | 14 (3.4) | 2.24 |  |  |
| *R. parilis* / 8805 | 3.633333333 | | 24 (3.6) | 2.37 |  |  |
| *R. polymorpha* / 8663o | 5.966666667 | | 25 (6.0) | 3.90 |  |  |
| *R. polymorpha* / 8663t | 4.966666667 | | 21 (5.0) | 3.24 |  |  |
| *R. polymorpha* / 8665u | 4.366666667 | | 28 (4.4) | 2.85 |  |  |
| *R. polymorpha* / 8668g | 7.266666667 | | 28 (7.3) | 4.75 |  |  |
| *R. polymorpha* / 8807\_3 | 2.933333333 | | 16 (2.9) | 1.92 |  |  |
| *R. polymorpha* / 8935b | 3.866666667 | | 19 (3.9) | 2.53 |  |  |
| *R. porteri* / 8663a | 5.1 | | 23 (5.1) | 3.33 |  |  |
| *R. porteri* / 8668a | 18.26666667 | | 27 (18.3) | 11.93 |  |  |
| *R. porteri* / 8795 | 3.966666667 | | 26 (4.0) | 2.59 |  |  |
| *R. porteri* / 8796 | 4.133333333 | | 43 (4.1) | 2.70 |  |  |
| *R. porteri* / 8797 | 4.9 | | 19 (4.9) | 3.20 |  |  |
| *R. shushanii* / 8664\_2 | 4.433333333 | | 22 (4.4) | 2.90 |  |  |
| *R. shushanii* / 8664\_3 | 4.566666667 | | 9 (4.6) | 2.98 |  |  |
| *R. shushanii* / 8664\_4 | 6.3 | | 16 (6.3) | 4.11 |  |  |
| *R. shushanii* / 8664\_5 | 4.6 | | 11 (4.6) | 3.00 |  |  |
| *R. shushanii* / 8664\_5 | 4.633333333 | | 29 (4.63) | 3.02 |  |  |
| *R. peltata* / 9101 | 13.16666667 | | 20 (13.2) | 8.60 |  |  |
| *R. subdiscrepan s*/ 9052 | 11.1 | | 29 (11.1) | 7.25 |  |  |
|  |  |  | |  |  |  |

**Fig. 1.** Variation

**Fig. 2**. Topologies. A. ITS; B. IGS; C. 28S; D. 28S introns; E. 18S; & F 18S introns.

DePriest, P.T., 1993. Molecular innovations in lichen systematics: the use of ribosomal and intron nucleotide sequences in the *Cladonia chlorophaea* complex. Bryologist, 314-325.

Divakar, P.K., Leavitt, S.D., Molina, M.C., Del-Prado, R., Lumbsch, H.T., Crespo, A., 2016. A DNA barcoding approach for identification of hidden diversity in Parmeliaceae (Ascomycota): *Parmelia sensu* stricto as a case study. Botanical Journal of the Linnean Society 180, 21-29.

Gargas, A., DePriest, P.T., Grube, M., Tehler, A., 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science 268, 1492-1495.

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.

Gutiérrez, G., Blanco, O., Divakar, P., Lumbsch, H., Crespo, A., 2007. Patterns of Group I Intron Presence in Nuclear SSU rDNA of the Lichen Family Parmeliaceae. Journal of Molecular Evolution 64, 181-195.

Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014. Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbriata* sensu stricto. Mycologia 106, 224-242.

Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. The Quarterly Review of Biology 66, 411-453.

Kelly, L.J., Hollingsworth, P.M., Coppins, B.J., Ellis, C.J., Harrold, P., Tosh, J., Yahr, R., 2011. DNA barcoding of lichenized fungi demonstrates high identification success in a floristic context. New Phytologist 191, 288-300.

Leavitt, S.D., Esslinger, T.L., Hansen, E.S., Divakar, P.K., Crespo, A., Loomis, B.F., Lumbsch, H.T., 2014. DNA barcoding of brown Parmeliae (Parmeliaceae) species: a molecular approach for accurate specimen identification, emphasizing species in Greenland. Organisms Diversity & Evolution 14, 11-20.

Leavitt, S.D., Fankhauser, J.D., Leavitt, D.H., Porter, L.D., Johnson, L.A., St. Clair, L.L., 2011. Complex patterns of speciation in cosmopolitan ‘‘rock posy’’ lichens – Discovering and delimiting cryptic fungal species in the lichen-forming *Rhizoplaca melanophthalma* species-complex (Lecanoraceae, Ascomycota). Molecular Phylogenetics and Evolution 59, 587-602.

Leavitt, S.D., Fernández-Mendoza, F., Pérez-Ortega, S., Sohrabi, M., Divakar, P.K., Lumbsch, H.T., St. Clair, L.L., 2013a. DNA barcode identification of lichen-forming fungal species in the *Rhizoplaca melanophthalma* species-complex (Lecanorales, Lecanoraceae), including five new species MycoKeys 7, 1–22.

Leavitt, S.D., Fernández-Mendoza, F., Pérez-Ortega, S., Sohrabi, M., Divakar, P.K., Vondrák, J., Thorsten Lumbsch, H., Clair, L.L.S., 2013b. Local representation of global diversity in a cosmopolitan lichen-forming fungal species complex (*Rhizoplaca*, Ascomycota). Journal of Biogeography 40, 1792–1806.

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

Lindner, D.L., Banik, M.T., 2011. Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. Mycologia 103, 731-740.

Lindner, D.L., Carlsen, T., Henrik Nilsson, R., Davey, M., Schumacher, T., Kauserud, H., 2013. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecology and Evolution 3, 1751-1764.

Lofgren, L.A., Uehling, J.K., Branco, S., Bruns, T.D., Martin, F., Kennedy, P.G., Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Molecular Ecology 0.

Myllys, L., Lohtander, K., Källersjö, M., Tehler, A., 1999. Sequence Insertions and ITS Data Provide Congruent Information on Roccella canariensis and R. tuberculata (Arthoniales, Euascomycetes) Phylogeny. Molecular phylogenetics and evolution 12, 295-309.

Nilsson, R.H., Larsson, K.-H., Taylor, A.F S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47, D259-D264.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Consortium, F.B., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences, doi: 10.1073/pnas.1117018109.

Szostak, J.W., Wu, R., 1980. Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature 284, 426-430.

Thell, A., 1999. Group I Intron Versus its Sequences in Phylogeny of Cetrarioid Lichens. The Lichenologist 31, 441-449.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, N., Gelfand, D., Sninsky, J., White, T.J. (Eds.), PCR protocols. Academic Press, San Diego, pp. 315-322.

Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S., Kristiansson, E., Kagami, M., Nilsson, R.H., Introducing ribosomal tandem repeat barcoding for fungi. Molecular Ecology Resources 0.