

# Characterizing the ribosomal tandem repeat and its utility as a DNA barcode in lichen-forming fungi

**Michael Bradshaw**

Brigham Young University

**Felix Grewe**

The Field Museum

**Anne Thomas**

Brigham Young University

**Cody H. Harrison**

Brigham Young University

**Hanna Lindgren**

The Field Museum

**Lucia Muggia**

Universita degli Studi di Trieste

**Larry L. St. Clair**

Brigham Young University

**H. Thorsten Lumbsch**

The Field Museum

**Steve Leavitt** (✉ [steve\\_leavitt@byu.edu](mailto:steve_leavitt@byu.edu))

Brigham Young University <https://orcid.org/0000-0002-5034-9724>

---

## Research article

**Keywords:** copy number variation, DNA barcoding, ITS, lichens, repeat region, Rhizoplaca

**Posted Date:** September 12th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.14347/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 6th, 2020. See the published version at <https://doi.org/10.1186/s12862-019-1571-4>.

# Abstract

**Background** 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 nuclear ribosomal DNA (nrDNA) evolution are poorly characterized for fungi in general, including lichenized fungi. The internal transcribed spacer (ITS) region of the nrDNA has been adopted as the primary DNA barcode identification marker for fungi. However, little is known about intragenomic variation in the nrDNA in symbiotic fungi. In order to better understand evolution of nrDNA and the utility of the ITS region for barcode identification of lichen-forming fungal species, we generated nearly complete nuclear ribosomal operon sequences from approximate nine species in the *Rhizoplaca melanophthalma* species complex using short reads from high-throughput sequencing.

**Results** We estimated copy numbers for the nrDNA operon, ranging from nine to 48 copies for members of this complex, and found low levels of intragenomic variation in the standard barcode region (ITS). 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.

**Conclusions** The results from our study support the general assumption that the ITS region of the nrDNA is an effective barcoding marker for fungi. For the *R. melanophthalma* group, the limited number of potentially polymorphic sites generally do not correspond to fixed diagnostic nucleotide position characters separating taxa within this species complex. Previously unrecognized lineages inferred from ITS sequence data may represent undescribed species-level lineages or reflect uncharacterized aspects of nrDNA evolution.

## Background

For eukaryotes, regions within the nuclear ribosomal (nrDNA) operon have been instrumental in characterizing diversity and inferring evolutionary relationships [1, 2]. 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 [3, 4]. Copy number of the nrDNA operon is a rapidly evolving trait [5]. Across Fungi, nrDNA copy number has been shown to vary considerably, ranging from tens to over 1400 copies per genome [6]. The relative ease of amplification, coupled with variable substitution rates among different regions of the nrDNA, have promoted its longstanding use in phylogenetic and biodiversity research [1]. Impetus for including the full nrDNA operon in published genome assemblies [7] and use of nrDNA in emerging long-read sequencing technologies [1] highlights the need of improved characterization of this important genomic region.

The internal transcribed spacer region—ITS: comprising ITS1, 5.8S, and ITS2—within the nrDNA operon has been proposed as the standard DNA barcoding region for Fungi [8]. The ITS barcode currently plays a fundamental role in characterizing fungal diversity [9]. 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 [7]. Some studies report that mutations and variation within the ITS region are relatively minor and of little practical consequence, attributing the consistency to concerted evolution [10]. Other studies, however, report significant amounts of variation, suggesting that evolution may not occur in a purely concerted manner [11]. Studies using traditional Sanger sequencing may effectively conceal potential intragenomic variation due to PCR bias or dominating signal from the predominantly amplified copy [12]. Cloned sequencing studies have revealed the occurrence of intragenomic variation in nrDNA in multiple fungal lineages [7, 13]. In fact, a pyrosequencing-based study suggests that distinct ITS copies are found in multiple Ascomycete and Basidiomycete lineages, although in a relatively low proportion of sampled lineages [14, 15].

Information from nrDNA has been fundamental in research into lichen-forming fungi [4, 16–19]. Although some studies have demonstrated limitations of the ITS barcoding for delimitation of species [20, 21], others highlight the potential of DNA barcoding studies to improve our ability to characterize diversity of lichen-forming fungi [22–26]. The full nrDNA operon, to our knowledge, has not yet been characterized for any of the lichenized fungi examined to date. Furthermore, in spite of suggestive evidence that some lineages of lichenized fungi harbor multiple distinct copies of the ITS region [27], intragenomic nrDNA variation has not yet been explicitly tested to our knowledge.

In this study we investigated members of the *Rhizoplaca melanophthalma* species complex [28] with a goal of more fully characterizing features of nrDNA in a defined group of lichen-forming Ascomycetes. The *Rhizoplaca melanophthalma* species complex is a monophyletic lineage currently consists of approximately ten closely-related species/species-level lineages that originated during the Miocene and diversified largely during and Pleistocene [29]. Previous empirical species delimitation studies have circumscribed robust species boundaries among closely-related and morphologically similar species [28, 30], a pattern which has been supported by genome-scale molecular data [31, 32]. All formally described species in the complex can be identified using the standard DNA barcoding marker [30], with the exception of the two vagrant species *R. haydenii* (Tuck.) W. A. Weber and *R. idahoensis* Rosentreter & McCune.

To more fully characterize the nuclear ribosomal operon for 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, 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 ITS region of the nrDNA operon as a barcoding marker for fungi.

## Result

Sequences and alignments generated and used in this study were submitted to TreeBase (<https://treebase.org/>; study No. 24225).

## 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 the 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, the 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') (Additional file 1).

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 considerably 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 regions. The aligned nrDNA operon (18S, ITS1, 5.8 S, ITS2, 28S, and IGS) included a total of 18,386 nucleotide position characters.

## 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 region—in members of the *R. melanophthalma* group, with variance at a single nucleotide position character rarely exceeding 10% (Additional file 2; Additional file 3). Furthermore, potentially polymorphic sites generally did not coincide with segregating sites that separated species (Additional file 3). 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 of nrDNA regions with single-copy regions of the nuclear genome (data not shown). 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 nrDNA copy number.

A total of 13 introns were identified within the 18S and ten in the 28S genes; and the intron occurrences of *Rhizoplaca* and the outgroup taxa was compared with the phylogeny inferred from the complete, aligned nrDNA region (Fig. 1).

## Phylogenetic relationships inferred from nrDNA

Topologies independently 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. 2E). 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 & 2F). 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. 1).

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. 3; Additional file 4). *Rhizoplaca polymorpha* was recovered as monophyletic with weak statistical support while *R. haydenii* was recovered in two separate clades, one corresponding to *R. haydenii* spp. *arbuscula* and the other comprised of specimens representing *R. haydenii* spp. *haydenii* and *R. idahoensis* (Fig. 3). Three previously undetected clades were recovered in the ITS/IGS topology—'nrDNA clade I', 'nrDNA clade II', and 'nrDNA clade III' (Fig. 3).

## Discussion

Here we provide evidence of intragenomic copy number variation of the nrDNA operon in the *Rhizoplaca melanophthalma* species group, ranging from nine to 43 copies. Our estimates for members of the *Rhizoplaca melanophthalma* species group were in line with recent genome-based estimates for symbiotrophic fungi [6]. Given the rather limited intragenomic variation among copies (Additional file 3), it appears that specimen identification using the standard fungal DNA barcoding marker, the ITS [8], is not biased by intragenomic variation in the nrDNA region for the *Rhizoplaca melanophthalma* species group [30]. While copy number variation in fungi has recently been investigated and discussed in light of our current understanding [6], below we highlight the implications of our findings as they relate to the utility of nrDNA for specimen identification, patterns of intron evolution, and the development of complete nrDNA reference libraries.

We used short reads from metagenomes of lichenized fungi to infer limited intragenomic variation in members of the *R. melanophthalma* species complex. Potentially variable sites among nrDNA operon copies did not coincide with diagnostic, fixed nucleotide position characters separating species in the *Rhizoplaca melanophthalma* species group (Additional file 3), providing additional evidence that distinct clades in the nrDNA topologies are not merely a reflection of variable nrDNA copies. Similarly, consistent clades are recovered from alignments of different regions of nrDNA, e.g., 28S, intergenic spacer region,

and intronic regions (Fig. 3) and distinct clades in the ITS are not merely idiosyncratic in the ITS region alone.

For members of the *R. melanophthalma* species complex, we note that the distinct clades recovered from phylogenetic analyses of nrDNA do not correspond to phylogenies reconstructed from genome-scale data, specifically for members of the *R. porteri* group—*R. occulta*, *R. polymorpha*, and *R. porteri* [31, 32]. *Rhizoplaca occulta*, *R. polymorpha*, and *R. porteri* are recovered as divergent lineages in analyses of nrDNA, none of which are each another's sister clade (Fig. 3), while genome-scale data consistently recover these three taxa as closely related and intermixed in a well-supported clade [31, 32]. This is in contrast to *R. haydenii*, *R. melanophthalma*, *R. novomexicana*, *R. parilis*, and *R. shushanii* that are recovered as distinct in phylogenies inferred from both nrDNA alignments and phylogenomic data matrices.

The origin of highly divergent nrDNA clades among closely related, or even conspecific, lineages remains enigmatic. The three previously undetected nrDNA clades recovered here within the *R. melanophthalma* complex, 'nrDNA clade I', 'nrDNA clade II', and 'nrDNA clade III' (Fig. 3), may represent previously unsampled species-level lineages, other members of the *R. porteri* clade with divergent nrDNA, or some other unexplained phenomenon. The role of hybridization/introgression has recently been proposed as an important mechanism in the diversification of the *Rhizoplaca melanophthalma* species group (Keuler et al. *in review*). It is possible that divergent nrDNA clades may represent evidence of reticulate evolution with extinct or as-of-yet unsampled species or simply artifacts of hybridization/introgression. However, additional broader sampling will be required to infer the relationship of these newly found divergent nrDNA lineages. Similarly, the number and diversity of lichen associated symbionts has recently been shown to be far more complex than initially thought [33]. These studies suggest that the dynamics of symbiotic genetics may be synergistic with intricate interactions that combine to support these unique and complicated symbiotic systems. However, with the use of additional high-throughput sequencing methods, including single molecule sequencing and/or long-read sequencing technologies it will be possible to more directly characterize intragenomic variation and copy number of nrDNA [1, 34, 35].

Here, intragenomic variation was observed to be low in the *Rhizoplaca melanophthalma* species group. Concerted evolution is considered to be responsible for the maintenance of sequence similarity among copies of rDNA repeats in the operon potentially leading to low intragenomic variation [36, 37]. Concerted evolution is driven by unequal crossing over and gene conversion and although these two processes both lead to the homogenization of rDNA repeat units their underlying mechanisms differ from each other [37, 38]. Gene conversion is caused by replacement of DNA sequence by another in a unidirectional manner during homologous recombination. Unequal crossing over, on the other hand, involves the misalignment of homologous chromosomes in meiosis or sister chromatids in mitosis followed by nonreciprocal transfer of DNA sequence from one chromosomal location to another resulting in a deletion in one of the chromosomes and a duplication in the other. While both these mechanisms lead to homogenization of rDNA repeat units, unequal crossing over also has the potential to affect the copy number of the repeat units [36, 39].

The results of this study revealed striking patterns of multiple intron gains and losses in the *R. melanophthalma* group. Eleven of the 23 introns found in the 18S and 28S nrDNA genes are present in all *Rhizoplaca* specimens (i1, i5, i9, i10, i11 in the 18S gene and i3, i5, i7 to i10 in the 28S gene), but the pattern of all other introns is highly variable indicating multiple gains and losses in the evolution of *Rhizoplaca*. When comparing the intron pattern with the phylogenetic tree, only the occurrence of three introns can be most parsimoniously explained by a single event, i. e. the loss i7 from the 18S gene in some *R. porteri*, the loss of i13 from the 18S gene in all *R. porteri*, and a gain of i4 in the 28S gene after the split of *R. novomexicana* and *R. melanophthalma* from the remaining *Rhizoplaca*. All other introns would require at least require two (i2, i3, i4, i6, i12 in 18S and i1, i2, i6 in 28S) or four losses (i8 in 18S) to reassemble the intron pattern. This intron pattern confirms that ntDNA introns are highly mobile and the dynamic variability of the nrDNA region [40].

## Results

The results from our study support the general assumption that the ITS region of the nrDNA is an effective barcoding marker for fungi. The recent development of general primers that allow for amplification of the complete ribosomal operon, in conjunction with PacBio and Nanopore sequencing technologies, have helped promote the development of more comprehensive nrDNA databases [1]. The nearly complete ribosomal operons assembled for this study rank among the first generated for lichen-forming fungi. These data span highly conservative nrDNA regions that are important for high-level taxonomic classification to highly variable regions that can serve for population-level inference. A number of questions relating nrDNA evolution remain unanswered, including processes driving concerted evolution [36], copy number variation, particularly in symbiotic fungi [6], and patterns of intron distribution [4, 40, 41].

## Methods

### Taxonomic sampling and data compilation

Our sampling included representatives of the nine formally recognized species within the *R. melanophthalma* species complex [30] and two outgroup taxa—*R. subdiscrepans* (Nyl.) R. Sant. and *Protoparmeliopsis peltata* (Ramond) Arup, Zhao Xin & Lumbsch (Additional file 1). For this study, we analyzed short-read metagenomic data from a total of 33 specimens, representing ten *Rhizoplaca* s. lat. species (Leavitt et al., 2016), including: *R. haydenii* (Tuck.) W. A. Weber (n = 2), *R. melanophthalma* (DC.) Leuckert (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 representatives 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 for a total of 496 specimens from the *R. melanophthalma* species complex collected from sites throughout western North America, the center of species diversity for this group [29]. For all new specimens, DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega), and for amplification and sequencing of the ITS marker followed previously described methods were used [28]. Newly generated sequences were combined with previously available nrDNA sequence data from the *Rhizoplaca melanophthalma* group (<https://treebase.org/>; study No. 19048).

## Short-read data, genome assembly and identification of the nuclear ribosomal operon

Short reads from 33 *Rhizoplaca* specimens reported in a previous study [31] were used for genome assembly to identify contigs containing the nuclear ribosomal operon. Full details of specimen preparation and sequencing are described in [31]. In short, libraries were prepared using the Illumina Nextera XT DNA library prep kit (product discontinued), then pooled and sequenced using a single lane on the Illumina HiSeq2000 platform, generating 100-bp paired-end reads with a 350-bp insert size. Four specimens were sequenced individually on the MiSeq platform (Illumina) generating 250-bp paired-end (PE) reads with a 550-bp insert size. 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. [42–46].

All paired-end (PE) reads were filtered using TRIMMOMATIC v0.33 [47] 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 [48] while 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. From each assembly, contigs containing the nrDNA were identified using a custom BLAST [49] search implemented in the program Geneious R11 [50] against available regions of the nrDNA, e.g., nuLSU, IGS, and ITS, generated from *Rhizoplaca melanophthalma s. lat.* specimens.

Some regions of the nrDNA operon are highly conserved across divergent lineages (e.g. 18S, 5.8S and portions of 28S subunit), and reads from non-target genomes (e.g., the photobiont, accessory fungi, etc.) may potentially bias the interpretation of intragenomic variation within *Rhizoplaca* species. Therefore, we used a *de novo* assembly approach for all nrDNA reads in order 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 the 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 using 'medium-low sensitivity' parameter. Resulting contigs were searched against NCBI's GenBank database using BLAST to identify non-target contigs, which were then excluded from further analysis.

## Assessing intragenomic variation of the ITS, inferring copy number of the ribosomal operon, and intron identification

Our assessment of potential intragenomic variation focused on the ITS region (ITS1, 5.8S, and ITS2)—the standard fungal DNA barcoding marker for specimen identification [8]. 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 [51]. 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. The Samtools v1.6 genomics utilities package [52] 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 then generated to identify the bases aligned with each position of the reference sequence which was visually confirmed using Geneious v11. A python script was used to identify 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 each nucleotide position character was divided by coverage at that location. When calculating percent variance, there was no effort was made to identify bases within a read that were sequencing error versus 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,  $\leq 0.1\%$  is achieved for  $\geq 75\text{--}85\%$  of base [53], while true intragenomic variation would exceed the error rate for Illumina sequencing.

To estimate the total copy number of the nrDNA operon, we compared average read depth coverage of nrDNA relative to coverage of putative single-copy regions of the nuclear genome [6]. For each specimen, reads were mapped back to their 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 [31], and three 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 of the copy number of the nrDNA operon.

In contrast to most other eukaryotic genomes, yeast genomes have few introns [54]. 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 distinguish 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 be present in all species within the *R. melanophthalma* group, except *R. porteri* [28]; and the absence of this intron serves as a diagnostic character in the description of this taxon [30]. However, PCR amplifications may not provide an accurate assessment of repetitive genomic regions due to PCR bias or an 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 with the Geneious v11 Read Mapper, using the “medium-low sensitivity/ fast” settings, iterated 5 times. To test if this group I intron might be absent in some copies of nrDNA found 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 intron using a custom script.

## 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 [55, 56], 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 nrDNA, 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 distinguished using phylogenetic reconstructions of the barcoding marker for fungi, the ITS region (Leavitt et al., 2011; Leavitt et al., 2013a). Here we also investigated the question as to whether or not species within this complex can be recovered as monophyletic using other regions of nrDNA for phylogenetic inference. We reconstructed phylogenies from different regions of the ribosomal operon: (i) 18S nrDNA, excluding introns; (ii) introns within the 18S region; (iii) 28S nrDNA, excluding introns; (iv) introns within the 28S region; (v) concatenated 18S and 28S nrDNA, 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 nrDNA, 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 individually from each of these regions individually using the program RAxML v8.2.2 [57]. 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

nrDNA 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 [58] with our newly sampled specimens, resulting in a total of 496 specimens characters (TreeBase study No. 24225; <https://treebase.org/>). 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 larger dataset using IQTree v1.6.3 [59], treating each region (IGS and ITS), with 1,000 ultra-fast bootstrap replicates [60] to assess nodal support, and the best-fit substitution model as predicted by ModelFinder [61].

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

Short reads are deposited in NCBI's Sequence Read Archive (SRA, project number pending); and assembled sequences were submitted to GenBank (accession numbers pending)

### **Competing interests**

The authors declare that they have no competing interests.

### **Funding**

This study was funded by the College of Life Sciences at Brigham Young University, The Grainger Bioinformatics Center, and the Negaunee Foundation.

### **Authors' contributions**

The idea of this study was conceived by SDL, FG, HL and HTL. MB, AT, CHH, FG and SDL generated and analyzed the data. MB, SDL and FG wrote the initial draft of the manuscript. All authors provided crucial conceptual feedback throughout the process and read and approved the final manuscript.

## Acknowledgements

We thank Todd Widhelm, Roger Rosentreter and Pradeep K. Divakar for valuable discussion and insight.

## References

1. Wurzbacher C, Larsson E, Bengtsson-Palme J, Van den Wyngaert S, Svantesson S, Kristiansson E, Kagami M, Nilsson RH: Introducing ribosomal tandem repeat barcoding for fungi. *Molecular Ecology Resources* 2019, 19(1):118-127.
2. White TJ, Bruns T, Lee S, Taylor J: Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols*. Edited by Innis N, Gelfand D, Sninsky J, White TJ. San Diego: Academic Press; 1990: 315-322.
3. Hillis DM, Dixon MT: Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *The Quarterly Review of Biology* 1991, 66(4):411-453.
4. Gutiérrez G, Blanco O, Divakar P, Lumbsch H, Crespo A: Patterns of Group I Intron Presence in Nuclear SSU rDNA of the Lichen Family Parmeliaceae. *Journal of Molecular Evolution* 2007, 64(2):181-195.
5. Szostak JW, Wu R: Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* 1980, 284(5755):426-430.
6. Lofgren LA, Uehling JK, Branco S, Bruns TD, Martin F, Kennedy PG: Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology* 2019, 28:721-730.
7. Lindner DL, Banik MT: Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia* 2011, 103(4):731-740.
8. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Consortium FB: Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* 2012:doi: 10.1073/pnas.1117018109.
9. Nilsson RH, Larsson K-H, Taylor AF S, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L *et al*: The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* 2019, 47(D1):D259-D264.

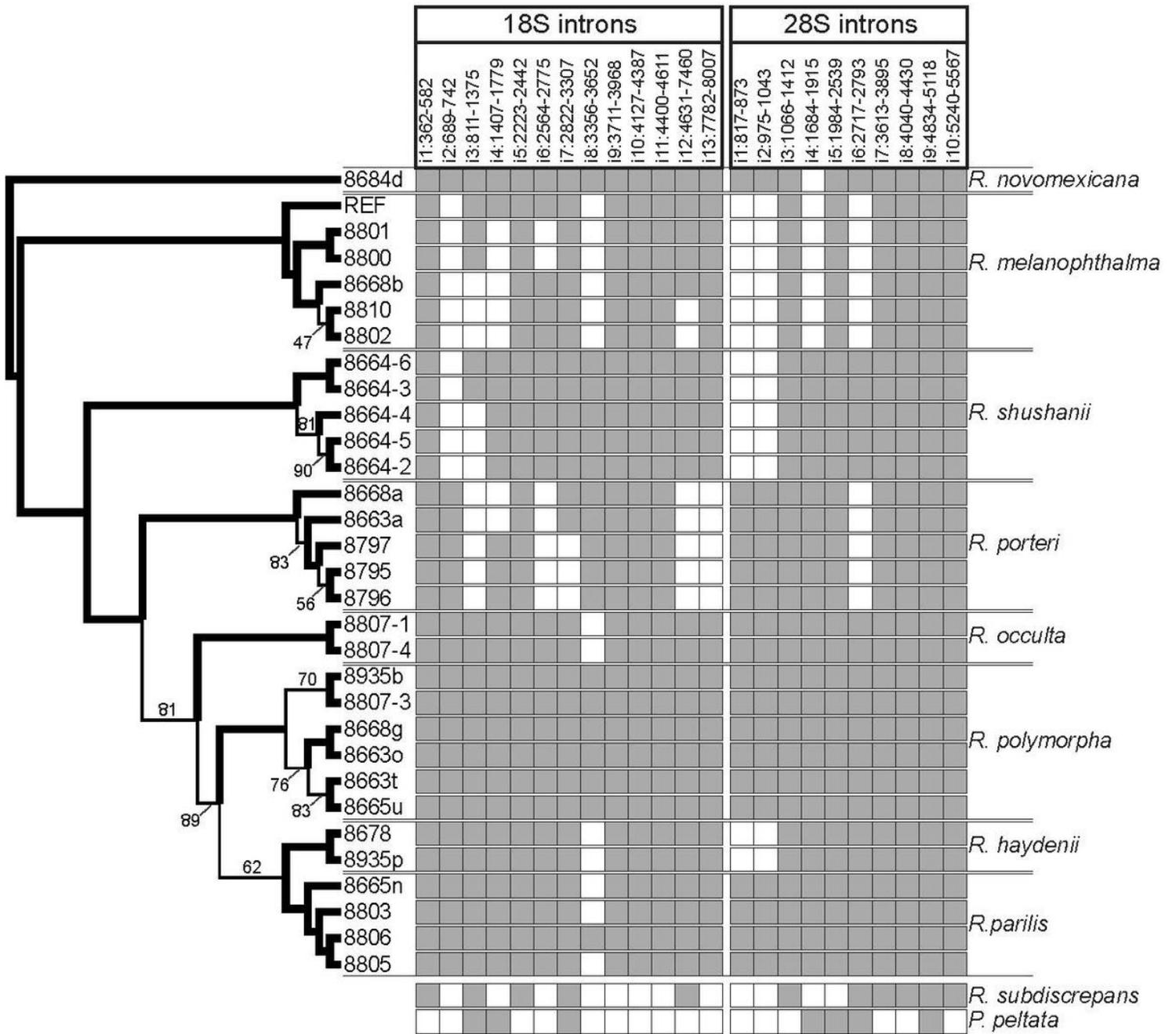
10. Ganley ARD, Kobayashi T: Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Research* 2007, 17(2):184-191.
11. Simon UK, Weiß M: Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution* 2008, 25(11):2251-2254.
12. Green SJ, Venkatramanan R, Naqib A: Deconstructing the polymerase chain reaction: Understanding and correcting bias associated with primer degeneracies and primer-template mismatches. *PlosOne* 2015, 10(5):e0128122.
13. Harrington TC, Kazmi MR, Al-Sadi AM, Ismail SI: Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbriata* sensu stricto. *Mycologia* 2014, 106(2):224-242.
14. Lindner DL, Carlsen T, Henrik Nilsson R, Davey M, Schumacher T, Kauserud H: Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution* 2013, 3(6):1751-1764.
15. Mark K, Cornejo C, Keller C, Flück D, Scheidegger C: Barcoding lichen-forming fungi using 454 pyrosequencing is challenged by artifactual and biological sequence variation. *Genome* 2016, 59(9):685-704.
16. Myllys L, Lohtander K, Källersjö M, Tehler A: Sequence Insertions and ITS Data Provide Congruent Information on *Roccella canariensis* and *R. tuberculata* (Arthoniales, Euascomycetes) Phylogeny. *Molecular phylogenetics and evolution* 1999, 12(3):295-309.
17. Thell A: Group I Intron Versus its Sequences in Phylogeny of Cetrarioid Lichens. *The Lichenologist* 1999, 31(05):441-449.
18. Gargas A, DePriest PT, Grube M, Tehler A: Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. *Science* 1995, 268(5216):1492-1495.
19. DePriest PT: Molecular innovations in lichen systematics: the use of ribosomal and intron nucleotide sequences in the *Cladonia chlorophaea* complex. *Bryologist* 1993:314-325.
20. Pino-Bodas R, Martin AP, Burgaz AR, Lumbsch HT: Species delimitation in *Cladonia* (Ascomycota): a challenge to the DNA barcoding philosophy. *Molecular Ecology Resources* 2013, in press.
21. Sadowska-Deś AD, Bálint M, Otte J, Schmitt I: Assessing intraspecific diversity in a lichen-forming fungus and its green algal symbiont: Evaluation of eight molecular markers. *Fungal Ecology* 2013, 6(2):141-151.
22. Leavitt SD, Esslinger TL, Hansen ES, Divakar PK, Crespo A, Loomis BF, Lumbsch HT: DNA barcoding of brown Parmeliae (Parmeliaceae) species: a molecular approach for accurate specimen identification, emphasizing species in Greenland. *Organisms Diversity & Evolution* 2014, 14(1):11-20.
23. Kelly LJ, Hollingsworth PM, Coppins BJ, Ellis CJ, Harrold P, Tosh J, Yahr R: DNA barcoding of lichenized fungi demonstrates high identification success in a floristic context. *New Phytologist* 2011, 191(1):288-300.

24. Divakar PK, Leavitt SD, Molina MC, Del-Prado R, Lumbsch HT, Crespo A: 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* 2016, 180(1):21-29.
25. Kanz B, Brackel Wv, Cezanne R, Eichler M, Hohmann M-L, Teuber D, Printzen C: DNA barcodes for the distinction of reindeer lichens: a case study using *Cladonia rangiferina* and *C. stygia*. *Herzogia* 2015, 28(2):445-464.
26. Yahr R, Schoch Conrad L, Dentinger Bryn TM: Scaling up discovery of hidden diversity in fungi: impacts of barcoding approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2016, 371(1702):20150336.
27. Simon DM, Hummel CL, Sheeley SL, Bhattacharya D: Heterogeneity of intron presence or absence in rDNA genes of the lichen species *Physcia aipolia* and *P. stellaris*. *Current Genetics* 2005, 47(6):389-399.
28. Leavitt SD, Fankhauser JD, Leavitt DH, Porter LD, Johnson LA, St. Clair LL: 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* 2011, 59(3):587-602.
29. Leavitt SD, Fernández-Mendoza F, Pérez-Ortega S, Sohrabi M, Divakar PK, Vondrák J, Thorsten Lumbsch H, Clair LLS: Local representation of global diversity in a cosmopolitan lichen-forming fungal species complex (*Rhizoplaca*, Ascomycota). *Journal of Biogeography* 2013, 40(9):1792-1806.
30. Leavitt SD, Fernández-Mendoza F, Pérez-Ortega S, Sohrabi M, Divakar PK, Lumbsch HT, St. Clair LL: DNA barcode identification of lichen-forming fungal species in the *Rhizoplaca melanophthalma* species-complex (Lecanorales, Lecanoraceae), including five new species *MycoKeys* 2013, 7:1-22.
31. Leavitt SD, Grewe F, Widhelm T, Muggia L, Wray B, Lumbsch HT: Resolving evolutionary relationships in lichen-forming fungi using diverse phylogenomic datasets and analytical approaches. *Scientific reports* 2016, 6:22262
32. Grewe F, Huang J-P, Leavitt SD, Lumbsch HT: Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. *Scientific Reports* 2017, 7(1):9884.
33. Spribille T: Relative symbiont input and the lichen symbiotic outcome. *Current Opinion in Plant Biology* 2018, 44:57-63.
34. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC *et al*: Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE* 2012, 7(11):e47768.
35. Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H: Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* 2009, 4:265.

36. Nei M, Rooney AP: Concerted and Birth-and-Death Evolution of Multigene Families. *Annual Review of Genetics* 2005, 39(1):121-152.
37. Eickbush TH, Eickbush DG: Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 2007, 175(2):477-485.
38. Naidoo K, Steenkamp ET, Coetzee MPA, Wingfield MJ, Wingfield BD: Concerted Evolution in the Ribosomal RNA Cistron. *PLOS ONE* 2013, 8(3):e59355.
39. Pinhal D, Yoshimura TS, Araki CS, Martins C: The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evolutionary Biology* 2011, 11(1):151.
40. DePriest PT, Been MD: Numerous group I introns with variable distributions in the ribosomal DNA of a lichen fungus. *Journal of Molecular Biology* 1992, 228(2):315-321.
41. Hibbett DS: Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. *Molecular Biology and Evolution* 1996, 13(7):903-917.
42. Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, Hofstetter V, Kauff F, Lutzoni F: A Phylogenetic Estimation of Trophic Transition Networks for Ascomycetous Fungi: Are Lichens Cradles of Symbiotrophic Fungal Diversification? *Systematic Biology* 2009, 58(3):283-297.
43. Cardinale M, Jr JvDc, Müller H, Berg G, Grube M: *In situ* analysis of the bacterial community associated with the reindeer lichen *Cladonia arbuscula* reveals predominance of *Alphaproteobacteria*. *FEMS Microbiology Ecology* 2008, 66(1):63-71.
44. Grube M, Cardinale M, de Castro JV, Jr., Muller H, Berg G: Species-specific structural and functional diversity of bacterial communities in lichen symbioses. *ISME J* 2009, 3(9):1105-1115.
45. Hodkinson B, Lutzoni F: A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. *Symbiosis* 2009, 49(3):163-180.
46. Muggia L, Vancurova L, Škaloud P, Peksa O, Wedin M, Grube M: The symbiotic playground of lichen thalli – a highly flexible photobiont association in rock-inhabiting lichens. *FEMS Microbiology Ecology* 2013, 85(2):313-323.
47. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30:2114-2120.
48. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y *et al*: Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: *Research in Computational Molecular Biology*. Edited by Deng M, Jiang R, Sun F, Zhang X, vol. 7821: Springer Berlin Heidelberg; 2013: 158-170.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *Journal of Molecular Biology* 1990, 215(3):403-410.

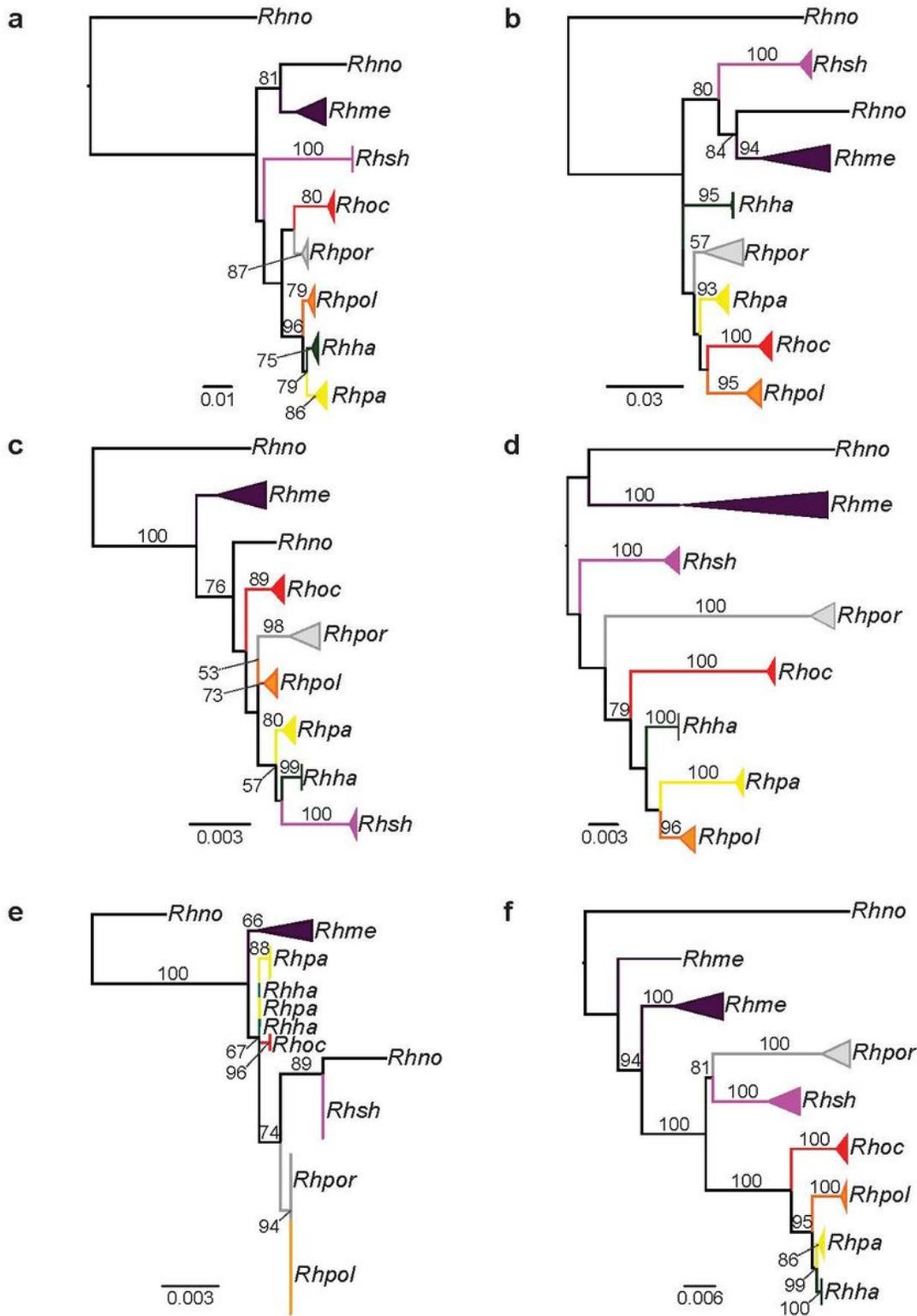
50. Kearsse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, 28(12):1647-1649.
51. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010, 26(5):589-595.
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPD: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25(16):2078-2079.
53. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. *Genome Biol* 2013, 14(5):R51.
54. Spingola M, Grate L, Haussler D, Ares M: Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 1999, 5(2):221-234.
55. Katoh K, Kuma K-i, Toh H, Miyata T: MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 2005, 33(2):511-518.
56. Katoh K, Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 2008, 9(4):286-298.
57. Stamatakis A: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014, 30(9):1312-1313.
58. Leavitt SD, Kraichak E, Vondrak J, Nelsen MP, Sohrabi M, Perez-Ortega S, St Clair LL, Lumbsch HT: Cryptic diversity and symbiont interactions in rock-psy lichens. *Molecular Phylogenetics and Evolution* 2016, 99:261-274.
59. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ: IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2014, 32(1):268-274.
60. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS: UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018, 35(2):518-522.
61. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS: ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017, 14(6):587-589.

## Figures



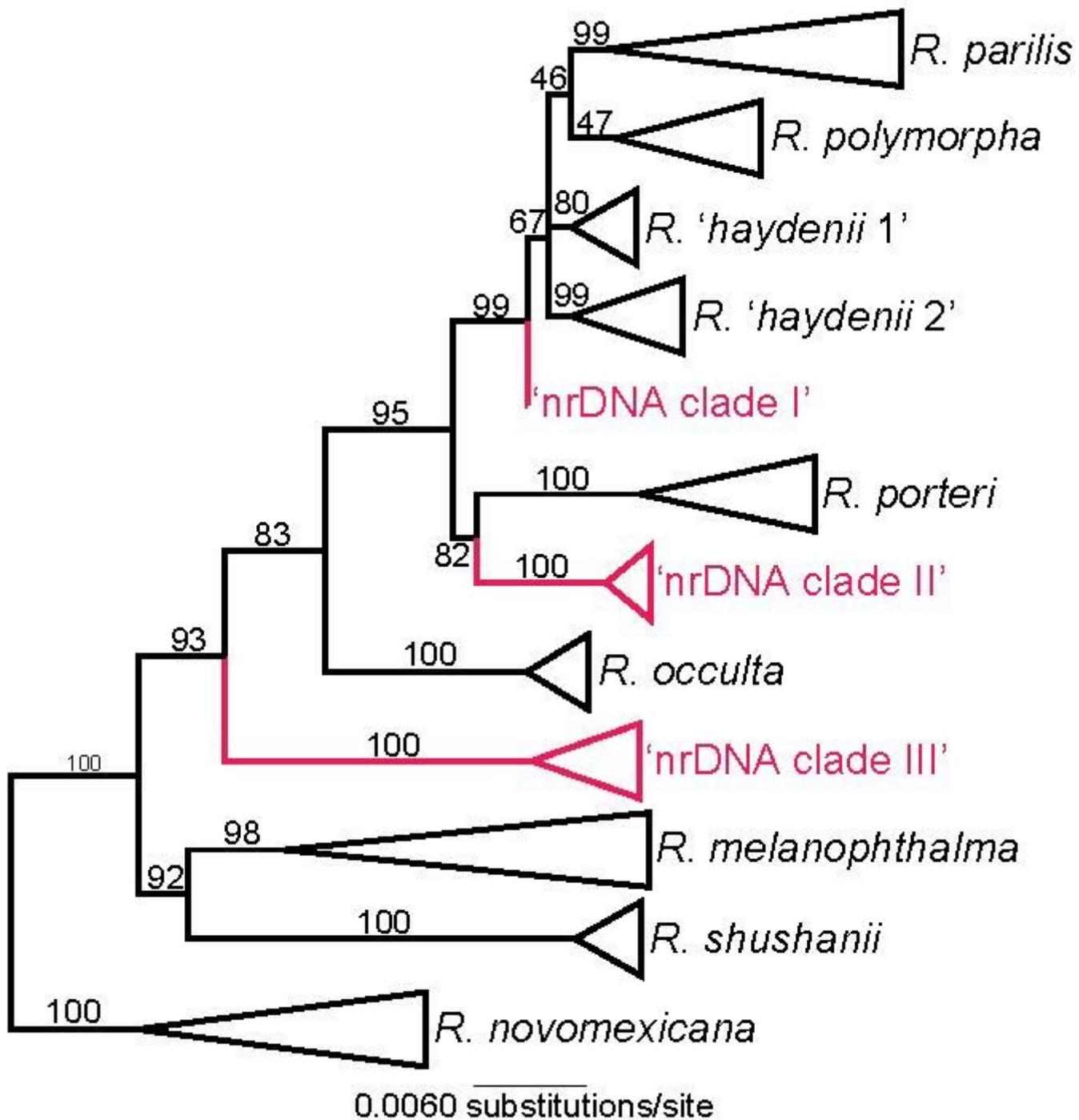
**Figure 1**

Phylogeny of the *Rhizoplaca melanophthalma* species complex inferred from the entire nuclear ribosomal cistron (18,386 nucleotide position characters) and intron patterns in the 18S and 28S regions. Bolded branches indicate bootstrap support values (BS)  $\geq 95\%$ , and BS  $< 95\%$  are indicated at nodes. The presence of introns – 13 introns in the 18S region and 10 in the 28S region – for each specimen is indicated by grey-filled squares. Presence and absence of introns is indicated for outgroups samples – *Protoparmeliopsis peltata* and *R. subdiscrepans* (phylogenetic relationships not shown).



**Figure 2**

Topologies inferred from various regions of the nuclear ribosomal cistron. a, topology inferred from the complete ITS region (ITS1f, 5.8S, and ITS2); b, topology inferred from the intergenic spacer region (IGS); c, topology inferred from the large subunit (28S); d, topology inferred from introns within the 28S region; e, topology inferred from the small subunit region (18S); & f, topology inferred from introns within the 18S region.



**Figure 3**

Simplified phylogeny inferred from a broad sampling (n=496) of concatenated ITS and IGS sequences. Bootstrap support values are indicated at nodes; and previously unknown nrDNA clades are shown in red text – “nrDNA clade I”, “nrDNA clade II”, and “nrDNA clade III” (online only). The complete phylogeny is provided as Additional file 4.