Received: 10 July 2018 | Revised: 22 November 2018 | Accepted: 27 November 2018

DOI: 10.1111/mec.14995

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles

Lotus A. Lofgren[1] | Jessie K. Uehling[2] | Sara Branco[3] | Thomas D. Bruns[2] | Francis Martin[4] | Peter G. Kennedy[1,5]

[1]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota

[2]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, California

[3]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana

[4]Laboratoire d'Excellence ARBRE, Interactions Arbres/Micro-organismes, INRA, UMR1136 INRA-Université de Lorraine, Champenoux, France

[5]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, Minnesota

Correspondence
Lotus A. Lofgren, Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN.
Email: LLofgren@umn.edu

Funding information
National Science Foundation, Grant/Award Number: 1554375

Abstract

Ribosomal DNA (rDNA) copy number variation (CNV) has major physiological implications for all organisms, but how it varies for fungi, an ecologically ubiquitous and important group of microorganisms, has yet to be systemically investigated. Here, we examine rDNA CNV using an in silico read depth approach for 91 fungal taxa with sequenced genomes and assess copy number conservation across phylogenetic scales and ecological lifestyles. rDNA copy number varied considerably across fungi, ranging from an estimated 14 to 1,442 copies (mean = 113, median = 82), and copy number similarity was inversely correlated with phylogenetic distance. No correlations were found between rDNA CNV and fungal trophic mode, ecological guild or genome size. Taken together, these results show that like other microorganisms, fungi exhibit substantial variation in rDNA copy number, which is linked to their phylogeny in a scale-dependent manner.

KEYWORDS
copy number variation, fungi, lifestyle, phylogenetic signal, rDNA, read depth

1 | INTRODUCTION

Ribosomes are a central component of life on Earth, and to meet varying needs for protein production, the genomes of most eukaryotic organisms contain multiple copies of ribosomal DNA (rDNA). There is considerable rDNA copy number variation (CNV) both within and among taxonomic groups, typically totalling less than 15 copies in prokaryotes (Liao, 2000), 39–19,300 copies in higher animals (Prokopowich, Gregory, & Crease, 2003), 150–26,048 copies in plants (Prokopowich et al., 2003) and up to 315,786 copies in ciliates (Gong, Dong, Liu, & Massana, 2013). Copy number is a rapidly evolving trait, and mechanisms for both rDNA copy number expansion and contraction have been described (Szostak & Wu, 1980). The consequences of rDNA CNV have received considerable attention in the context of DNA damage response (Ide, Miyazaki, Maki, & Kobayashi, 2010), DNA replication stress (Salim et al., 2017) and the expression of nonribosomal genes (Paredes, Branco, Hartl, Maggert, & Lemos, 2011). Similarly, the ecological importance of rDNA CNV has also been well characterized, with rDNA copy number being linked to ecosystem stoichiometry (Elser et al., 2000), growth rate and competitive ability (Klappenbach, Dunbar, & Schmidt, 2000; Nemergut et al., 2016) as well as bias in estimates of organismal abundance in high-throughput amplicon sequencing (Kembel et al., 2012, Perisin, Vetter, Gilbert, & Bergelson, 2016).

Relative to other microorganisms, estimates of rDNA CNV for fungi have been limited, and consequently, there has been no large-scale analysis of rDNA CNV with this ecologically important group of microorganisms. From the studies available, fungal rDNA CNV has been estimated to range between 28 and 511 (Liti et al., 2009;

Maleszka & Clarkwalker, 1993), which falls intermediate between prokaryotes and many larger eukaryotes. There have also been estimates of considerable rDNA CNV among strains of the same fungal species, with Liti et al. (2009) estimating that different strains of *Saccharomyces cerevisiae* had rDNA copies ranging from 54 to 511. The dikaryotic nature of many fungi suggests there may even be

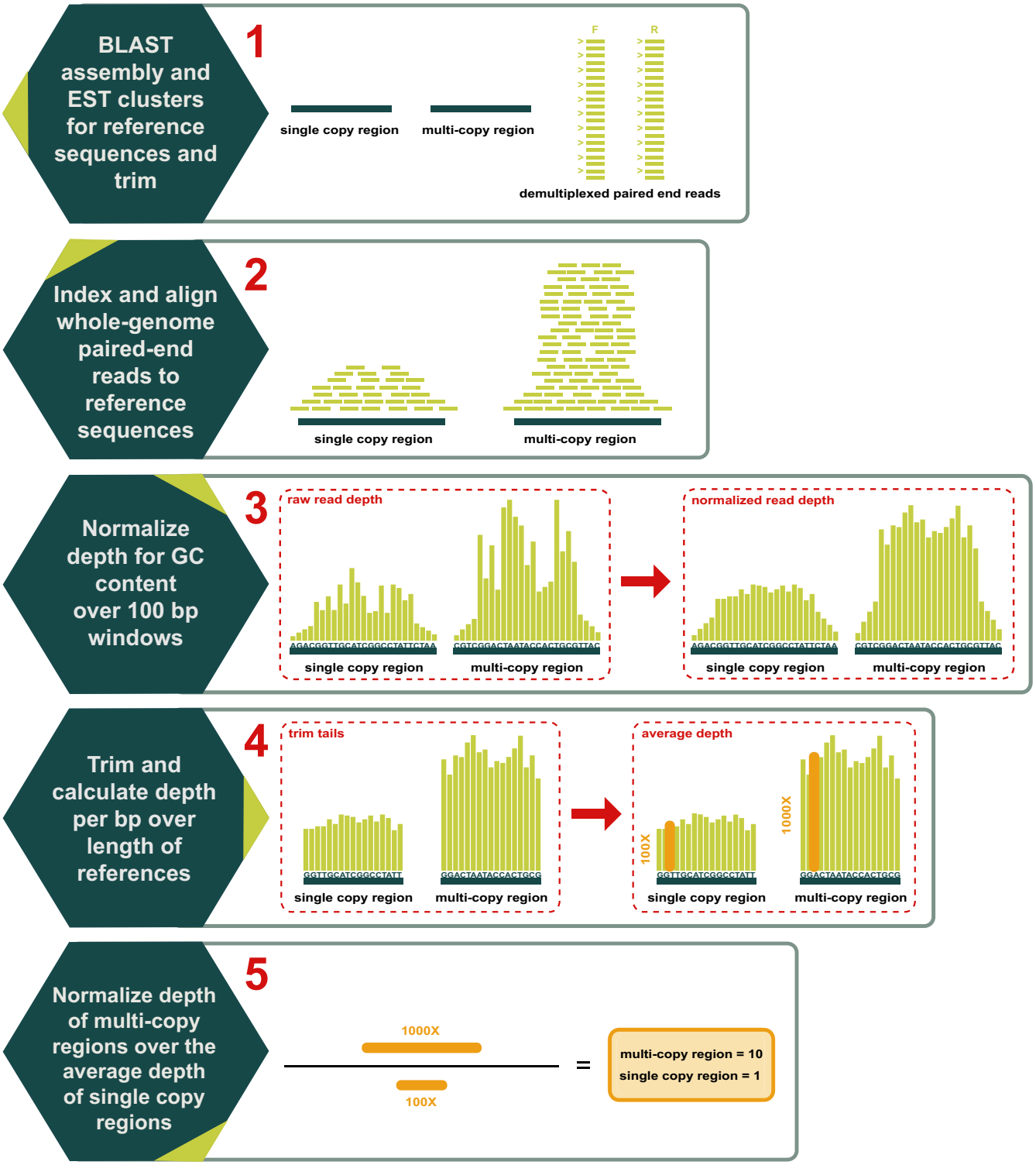

FIGURE 1 Analysis pipeline for estimating rDNA copy number. (1) Demultiplexed paired-end reads from whole-genome sequencing projects, along with 10 single-copy reference genes, are collected for each species. (2) Paired-end reads aligned to indexed references using Bowtie2. (3) Alignment depth over each reference bp is normalized for variable GC content using a 100-bp sliding window. (4) Overhangs are trimmed from alignments and average depth calculated over the length of each reference. (5) Number of rDNA copies is calculated by dividing the average depth of single-copy alignments by the average depth of multi-copy alignments (ITS and LSU) [Colour figure can be viewed at wileyonlinelibrary.com]

rDNA CNV among genetically distinct nuclei within a fungal individual (Zolan, 1995).

Despite a rapid increase in the sequencing of fungal genomes in recent years, estimates of rDNA CNV from annotated genomes have remained hindered by the collapsing of repetitive regions into a single representation. One solution to this problem is comparing the abundance of raw reads aligned to both single and multi-copy regions of DNA, an approach commonly known as relative read depth. Analysis of CNV using read depth was first developed to analyse repeat variation in tumour genomes (Chiang et al., 2009), and later used to account for anomalies in 16S read abundance in bacteria (Perisin et al., 2016). Here, we apply this approach to estimate rDNA copy number across a phylogenetically and ecologically diverse suite of fungi (Figure 1).

Based on the significant positive relationships observed between rDNA copy number and phylogenetic relatedness among other microorganisms (Kembel et al., 2012), we hypothesized that variation in rDNA copy number would exhibit significant phylogenetic signal in fungi. Additionally, due to the association of rDNA copy number and the physiological phenomena noted above, we hypothesized that rDNA CNV would also be linked with fungal ecological lifestyle. Specifically, because rDNA-associated traits such as rapid growth or stress tolerance may be more crucial for some fungal lifestyles than others (e.g., pathogens vs. mutualist fungi), we predicted that there would be a significant association between fungal ecological lifestyle and rDNA copy number. Finally, because rDNA copy number has been reported to be significantly positively correlated with genome size in other eukaryotes (Prokopowich et al., 2003), we investigated the relationship between rDNA copy number and genome size, both dependent and independent of size contributions from rDNA in each genome.

2 | MATERIALS AND METHODS

2.1 | Copy number estimation pipeline

To assess rDNA CNV across a broad phylogenetic range of fungi, we selected 91 taxa with available genomic data, spanning phyla to interspecific populations. We also choose isolates to represent a wide variety of ecological lifestyles, including pathogens, saprotrophs, plant mutualists and taxa capable of multiple lifestyles. Raw reads for each taxon were transferred from the Joint Genome Institute's MycoCosm site (Grigoriev et al., 2014) to server space at the Minnesota Supercomputing Institute (MSI) using Globus (Foster, 2006). Quality scores were converted to PHRED33 using Trimmomatic where necessary (Bolger, Lohse, & Usadel, 2014). The ITS and LSU gene regions as well as 10 single-copy reference genes were collected for each sequenced taxon. Single-copy genes were obtained as genomic.fasta files (with introns included) by keyword searching MycoCosm within the complete annotated assembly of each genome (Supporting information Table S1). Current sequencing technologies (including long-read platforms) do not produce reads long enough to span multiple copies of the full rDNA cassette. As such, reads from multi-copy regions, such as the internal transcribed spacer region (ITS) or the large subunit rRNA gene (LSU), are often unable to accrue the confidence values necessary to warrant placement and are therefore typically excluded from genome assemblies. To overcome this issue, we procured the ITS and LSU reference regions unique to each genome from the EST clusters associated with each sequencing project. This was accomplished by BLAST searching ITS and LSU ($E = 1.0 \times 10^{-5}$, word size = 11) sequences from the same genus (search sequence randomly chosen from NCBI) against the EST database associated with each genome on MycoCosm (Figure 1.1). The nucleotide sequences of these EST clusters, internal to each genome, were then used in all downstream analyses. To confirm that EST clusters were high-fidelity sequence representatives, we compared EST-derived ITS sequences with Sanger-sequenced ITS regions for a subset of the same strains ($n = 7$) that were used to generate the assemblies and found the average number of incongruences to be 1.2 bp. DNA for Sanger sequencing was extracted using the REDExtract-*N*-Amp plant kit (Sigma-Aldrich), followed by PCR amplification using the primer pair ITS1-F/ITS4 (Gardes & Bruns, 1993; White, Bruns, Lee, & Taylor, 1990) and sequenced at the University of Minnesota Genomics Center. Sequences were aligned using SEQUENCHER version 5.1 (Gene Codes Corporation, Ann Arbor, MI) using default parameters to count incongruences between EST- and Sanger-derived sequences. For taxa where JGI annotations (single-copy genes) or EST clusters (multi-copy genes) were not available, reference sequences were procured from raw reads using -mpileup from BCFTOOLS in the SAMTOOLS package (Li et al., 2009). ITS reference sequences were trimmed on either side of the priming regions for ITS1-F and ITS4 (Gardes & Bruns, 1993; White et al., 1990) leaving ITS reference regions that were approximately 650 bp in length. LSU reference sequences were trimmed at the priming region for LROR (Rehner & Samuels, 1995) and again 750 bp downstream.

Reference sequences were indexed using BOWTIE2 (Langmead & Salzberg, 2012). Demultiplexed, paired- end reads for each genome were aligned to each reference gene individually [parameters: paired-end and –very-sensitive-local mode with a maximum number of unknown base calls equal to 0.15 × read length, and alignment score benefits dependent on PHRED values] (Figure 1.2). Sorting and depth calculations were carried out using SAMTOOLS version 1.3 (Li et al., 2009), with an increased max depth of 1 Mbp, and excluding reads with average quality scores <20. To correct for GC bias, GC normalization was conducted using a custom R script (R Core Team, 2017), employing a sliding window method as conceived by Yoon, Xuan, Makarov, Ye, and Sebat (2009) (see script: gc_norm.R) (Figure 1.3). Depth was then averaged over the length of each gene, minus the first and last 50 bp (which had misrepresentatively low depth due to alignment overhangs) (Figure 1.4). For the 40% of taxa where such data were available, we analysed sequences generated across two independent sequencing lanes to estimate stochastic variation introduced during the sequencing process. Single-copy genes with an average depth outside one standard deviation of the median value for each independently sequenced lane were excluded from the analysis. The copy number for multi-copy regions was

estimated by dividing the GC normalized depth of the average depth of ITS and LSU by an average of the GC normalized depth across all single-copy regions (Figure 1.5), and averaged against the two independently sequenced lanes (where possible). All analyses were carried out using batch submission to the MSI computing cluster (see cnv_pipeline.pbs for pipeline bash script).

2.2 | *In silico* verification of copy number estimation pipeline

A mock genome was generated consisting of 52 million randomly drawn base pairs (which falls within the genome size range of the fungal taxa included) in R. Using the reference regions for *Suillus brevipes* (a randomly chosen reference species), 60 concatenated multi-copy cassettes consisting of tandem ITS and LSU repeats, along with the 10 single-copy reference genes for *S. brevipes*, were inserted into known, nonoverlapping locations in the mock genome (see script: generate_mock_genome.R). Twenty-seven independently drawn sets of paired-end reads were then generated, varying in size from 1 to 50 million reads, formatted as.fastq files with idealized quality scores of ~ (representing the highest possible PHRED value in ASCII code) and run through the ITS CNV pipeline (see script: generate_mock_reads.R) (Figure 2).

2.3 | Phylogenetic analysis

A phylogeny containing the 91 fungal taxa was constructed using DNA sequences from three single-copy genes: TOP2, GH63 and MCM7. Alignments for each gene were carried out using MUSCLE (Edgar, 2004) on the CIPRES portal (Miller *et al*. 2010) and trimmed using trimAl (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) to remove gaps and noninformative positions. Sequences from the three genes were then manually concatenated, realigned and retrimmed resulting in 8,096 informative positions. Phylogenetic analysis was conducted using RAxML HPC2 (Stamatakis, 2006) on XSEDE (Towns et al., 2014) run with default parameters, which utilized a 16 state GTR model and calculated bootstrap support based on 1,000 iterations. Results were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree).

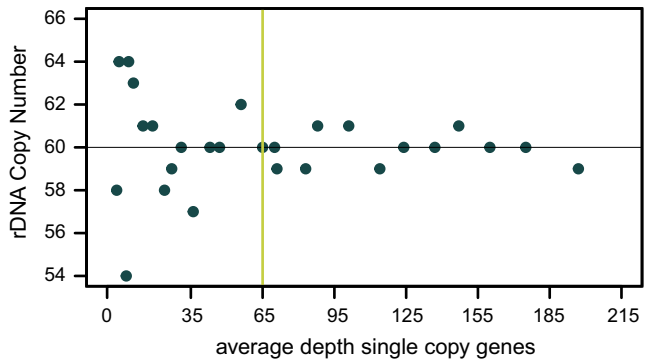

FIGURE 2 Mock genome pipeline validation. Observed rDNA copy number estimates for a mock genome containing 60 rDNA copies. Black horizontal line at 60 represents expected number of copies. Green vertical line indicates where copy number estimates are ±1 copy from expected after a depth of ~65×/bp [Colour figure can be viewed at wileyonlinelibrary.com]

2.4 | Statistical analyses

To determine whether fungal rDNA CNV displayed phylogenetic signal (i.e., conservation of rDNA copy number among more closely related taxa), we used the R package "phylosignal" (Keck et al., 2016) on the nonultrametric tree described above. This package calculates multiple evolutionary- and correlation-based metrics and allows for tests within internal nodes to identify significant "local hotspots" of trait conservation. Based on the recommendations of Münkemüller et al. (2012), Bloomberg's K and Pagel's λ were selected for the evolution-based metrics, while Abouheif's C_{mean} and Moran's I were selected for the spatial correlation metrics. The assessment of phylogenetic signal at internal nodes was conducted using the "lipaMoran" function, which calculates local Moran's I (I_i). To determine whether ecological lifestyle and rDNA CNV are related, we first assessed fungal taxa grouped by trophic mode—saprotrophic, pathotrophic, symbiotrophic, as well as those belonging to multiple trophic modes. We also tested differences in rDNA CNV among specific guilds containing sufficient taxon replication ($N \geq 5$): soil/litter/organic matter saprotroph versus pathogen within the Ascomycota and wood rot saprotroph versus ectomycorrhizal with the Basidiomycota. By delineating these latter two analyses by phyla, we sought to minimize the effect of phylogenetic relatedness (see results below). To determine significance in these ecological analyses, we used either parametric (ANOVA) or nonparametric (Kruskal–Wallis) tests depending on variance heterogeneity. Given the highly divergent rDNA copy number estimate for *Basidiobolus meristosporus* relative to all other taxa (see results below), we took a conservative approach and performed all the phylogenetic and ecological analyses with this taxon excluded.

For all genomes where full assembly sizes were published ($n = 79$), we analysed the correlation between rDNA CNV and genome size. Because repeat regions are not included in genome size estimates based on assembly size, we analysed the relationship between rDNA copy number and genome size both including and excluding the length contribution of the rDNA cassettes themselves. Length contribution from rDNA for each genome was estimated by taking the number of rDNA copies estimated for each genome and multiplying that by an assumed average rDNA cassette length of 9.1 kb (Miyazaki & Kobayashi, 2011) and then adding that additional size to each assembly size. Because phylogenetic signal analysis showed that rDNA CNV differed significantly by phyla (see below), we conducted correlational tests on both the whole data set and when the data set was subset by phylum. We used both parametric (Person's *r*) and nonparametric tests (Kendall's tau and Spearman's rho). For Pearson's *r*, the data were log-transformed when appropriate to normalize the distribution, according to visual inspection (Plotting) and numeric evidence (Shapiro–Wilk test of normality).

To account for phylogenetic autocorrelation, we also constructed a phylogenetic generalized least squares model (PGLS) with a subset tree constructed as above, and implemented in the R package Caper (Orme et al., 2014).

3 | RESULTS

From the 27 independent mock genome read libraries simulating variable sequencing depths, we found that our CNV estimation pipeline consistently returned the number of copies expected (±1 copy) after a read depth of 65× (Figure 2) (see scripts: generate_mock_genome.R, and generate_mock_reads.pbs). As such, we used 65X as the minimum read depth necessary to confidently estimate rDNA CNV. The estimates of rDNA copy number among the 91 taxa analysed exhibited some variation between sequencing lanes, with an average between lane difference of 14.9% (±2.4% SE) (Supporting information Table S2). The upper (251) and lower (11) limits of rDNA copy number estimates fell within the range of previous estimates for fungi, with the exception of *Basidiobolus meristosporus*, which had an estimated 1,442 rDNA copies (across fungi mean = 113 copies (98 with no outlier), median = 82 copies (with or without outlier), Figure 3a). Both the evolutionary (Bloomberg's K and Pagel's λ) and spatial correlation (Abouheif's C_{mean} and Moran's I) metrics indicated

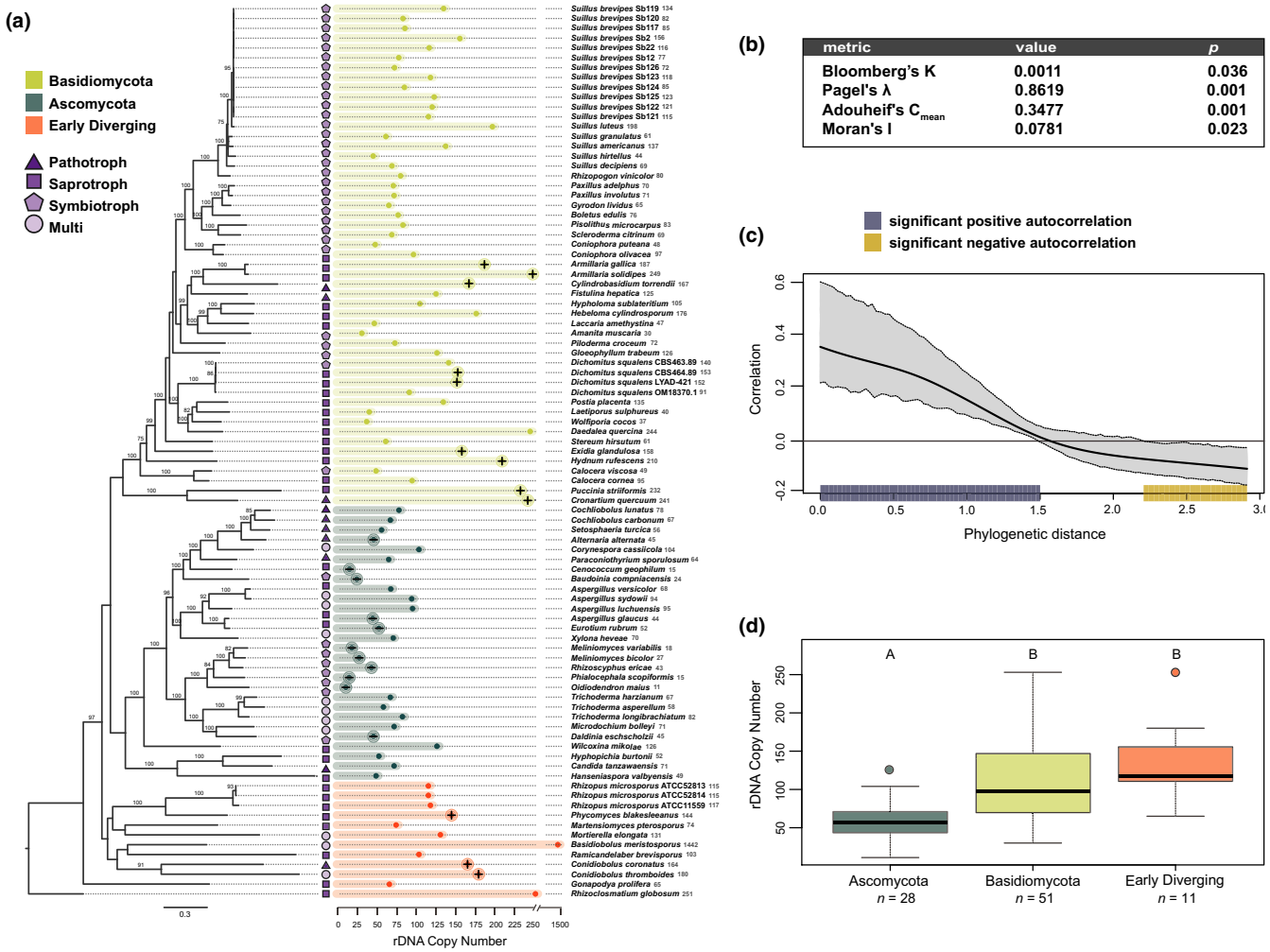

FIGURE 3 rDNA copy number variation across multiple phylogenetic scales. (a) A maximum-likelihood phylogenetic reconstruction of the 91 fungal taxa included in this study based on concatenation of three single-copy genes (TOP2, GH63 and MCM7). Branch values represent % bootstrap support from 1,000 iterations. Grey numbers next to taxa names indicate rDNA copy number. Ending targets on the copy number scale indicate values that are significantly positive (+) or significantly negative (–) according to local Moran's I, and highlight local hotspots of autocorrelation. (b) Significance tests of phylogenetic signal in rDNA copy number using both evolutionary (Bloomberg and Pagel) and autocorrelation (Abouheif and Moran) metrics. (c) Phylogenetic correlogram of autocorrelation based on Moran's I. The *x*-axis represents the patristic distance (unitless) of all pairwise comparisons for all taxa under investigation. Shaded area indicates the 95% confidence interval of autocorrelation values. Significance based on comparison to the null hypothesis of zero phylogenetic autocorrelation (horizontal black line at 0). (d) Distribution of rDNA copy number by fungal phylum. Different letters above groups indicate significant differences. Variance assumptions evaluated by Cochran's C test, and significance values assessed by ANOVA and Tukey HSD. See Supporting information Figure S1 for a validation that the observed differences in average copy number at the phylum level are not caused by overrepresentation of specific taxa [Colour figure can be viewed at wileyonlinelibrary.com]

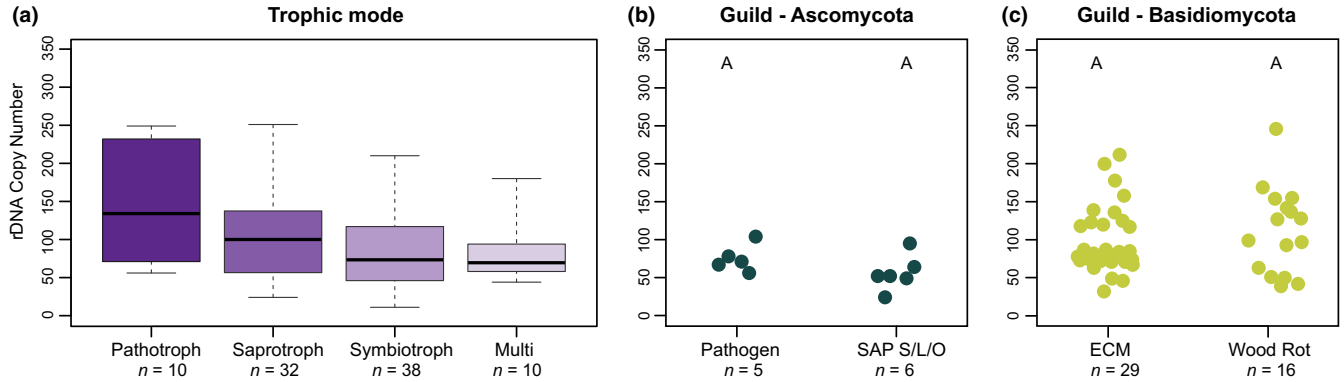

FIGURE 4 rDNA copy number variation by fungal ecological lifestyle. (a) Boxplot summaries of rDNA copy number variation by fungal trophic mode across the 91 taxa included. (b) rDNA copy number variation of Ascomycota taxa assigned to a specific fungal guild (sensu Nguyen et al, 2016); SAP S/L/O = soil, litter and organic matter saprophyte. (c) rDNA copy number variation of Basidiomycota taxa assigned to a specific fungal guild; ECM = ectomycorrhizal. For both b and c, only guilds with $n \geq 5$ replicate taxa were assessed. Variance assumptions evaluated by Cochran's C test and significance values assessed by Kruskal–Wallis (a) or ANOVA (b and c) tests; no significant differences were found [Colour figure can be viewed at wileyonlinelibrary.com]

significant phylogenetic signal in rDNA CNV (Figure 3b). Across the entire fungal phylogeny, there was a significant positive correlation between rDNA copy number and taxa at closer phylogenetic distances, but a significant negative correlation at greater distances (Figure 3c). The negative correlation was particularly notable at the level of phylum, where, on average, Ascomycota taxa had only half as many copies as those belonging to the Basidiomycota or early diverging lineages (Figure 3d). A similar trend in phylogenetic signal was also observed in the local Moran's index analyses, where all taxa with significant negative I_i values (rDNA copy number lower than expected) were in the Ascomycota and all those with significant positive I_i values (rDNA copy number higher than expected) belonged to the Basidiomycota or early diverging lineages (Figure 3a). With respect to ecological lifestyle, there were no significant differences in rDNA CNV across the three trophic modes or for taxa capable of belonging to multiple trophic modes (Figure 4a, Supporting information Table S3). When comparing among specific guilds, rDNA CNV was also not significantly different between pathogens and soil/litter/organic matter saprotrophs in the Ascomycota (Figure 4b) or between wood saprotrophs and ectomycorrhizal fungi in the Basidiomycota (Figure 4c). All tests examining the relationship between rDNA CNV and genome size failed to produce evidence that these metrics were correlated, regardless of the statistic used or the contribution of rDNA cassette length to total genome size (Figure 5).

4 | DISCUSSION

Our results indicate that rDNA CNV and phylogenetic relatedness are linked in fungi, but that this relationship is scale-dependent. At close phylogenetic scales (i.e., within species and genera), there was an overall trend of greater similarity in rDNA copy number, while at more distant scales (i.e., phyla), rDNA copy number became more divergent. Despite this general trend, we observed multiple examples that warrant caution when considering how rRNA copy number varies among fungi. For example, among the 12 different isolates of *Suillus brevipes*, estimates ranged from 72 to 156 copies, while across the genus (five additional species of *Suillus*), the estimated range was only slightly greater (44 to 198 copies). Interestingly, even at the very closest phylogenetic scale, CBS464.89 and CBS463.89 of *Dichomitus squalens*, which represent independent monokaryons from the same dikaryotic individual, had an estimated copy number difference of 13 (140 vs. 153). Although our analyses do confirm that total rDNA copy number is generally an order of magnitude greater for fungi than for bacteria or archaea, all three of these microbial groups display similar levels of variance in rDNA copy number (Větrovský & Baldrian 2013; Stoddard, Smith, Hein, Roller, & Schmidt, 2015). One notable exception to this trend was *B. meristosporus*. Because this species represents only a single isolate and a single sequencing library, this estimate should also be interpreted cautiously. However, *Basidiobolus spp.* have several properties that are unique, including a noncanonical nucleus-associated organelle, markedly large nuclei, and a genome that appears to be highly prone to duplication events (Henk & Fisher, 2012; McKerracher & Heath, 1985). Although there is evidence for the regulation of rDNA copy number, and some of the genetic mechanisms for rDNA copy number maintenance have been identified, (Kobayashi, Heck, Nomura, & Horiuchi, 1998; Russell & Rodland, 1986; Szostak & Wu, 1980), high rDNA variants have been reported in plants, animals and yeast (Rogers & Bendich 1987; Liti et al., 2009; Long et al., 2013) and it is currently unknown whether high rDNA CN strains represent a conserved or a temporary state (Pukkila & Skrzynia, 1993; Simon et al., 2018). Looking forward, research focused at multiple phylogenetic scales (e.g., is the amount of rDNA CNV observed within the genus *Suillus* common or an exception? Why are taxa in the Ascomycota consistently lower in rDNA copy number than other phyla?) represents important directions of study.

Although rDNA copy number is thought to have important physiological implications, such as allowing for more rapid growth (Stevenson & Schmidt 2004) and increased DNA damage response

(a) Genome size independent of rDNA contribution

metric	value	p
Pearson's pm	1.4797	0.147
Kendall's tau	0.6759	0.499
Spearman's rho	9529.9	0.515
Pearson's pm	–0.4199	0.678
Kendall's tau	–0.7713	0.441
Spearman's rho	4401.5	0.296
Pearson's pm	1.4071	0.193
Kendall's tau	1.2494	0.212
Spearman's rho	133.80	0.233
PGLS R^2	0.0021	0.7641

(b) Genome size with rDNA contribution

metric	value	p
Pearson's pm	1.7421	0.090
Kendall's tau	0.9089	0.363
Spearman's rho	9083.9	0.363
Pearson's pm	–0.3778	0.709
Kendall's tau	–0.6526	0.514
Spearman's rho	4346.5	0.334
Pearson's pm	1.5055	0.167
Kendall's tau	1.4056	0.160
Spearman's rho	122.78	0.174
PGLS R^2	0.0047	0.5492

(c)

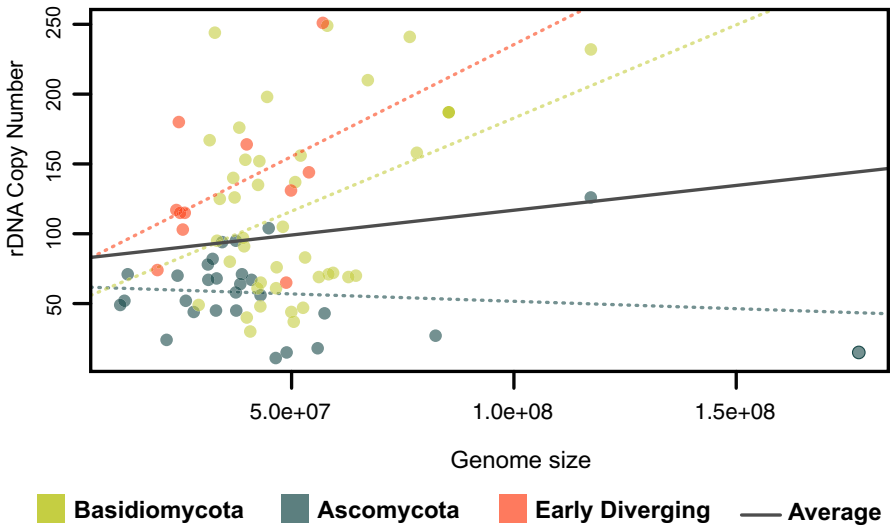

FIGURE 5 Fungal genome size and rDNA copy number variation. (a) Results based on genome assembly size, without including the length added by the rDNA cassettes. (b) Results after accounting for length added by rDNA cassettes. (c) Grey solid line represents the average across all taxa included, while dotted lines correspond to the rRNA CNV–genome size relationship for specific phyla. Relationships displayed are based on genome size without length contribution of rDNA cassettes included [Colour figure can be viewed at wileyonlinelibrary.com]

(Ide et al., 2010), our results did not find that rDNA copy number is coupled with fungal ecological lifestyle. Pathogenic fungi had a nonsignificant trend towards higher rDNA copy number in comparisons across trophic modes and between guilds, but there was considerable variation within this ecological lifestyle. Further, while genomic studies of fungi capable of using multiple trophic modes (e.g., saprotroph and symbiotroph) indicate that gene content and expression differs from taxa using a single trophic mode (Martino et al., 2018), we did not find evidence that this increased metabolic flexibility was correlated with rDNA copy number. Finally, within the Basidiomycota, ectomycorrhizal fungal representatives had rDNA copy number estimates that were very comparable to their saprotrophic wood rot relatives (Kohler et al., 2015). Given that previous studies have shown positive associations between rDNA copy number and traits relevant to lifestyle (Stevenson & Schmidt 2004; Ide et al., 2010), we suspect that the relatively coarse ecological scale of our analyses was not sufficient to capture clear links to fungal lifestyle. These results do, however, have notable ecological implications for estimates of fungal species abundances in high-throughput amplicon sequencing data

sets (Baldrian et al., 2012). Systematic bias may be introduced as a consequence of CNV-associated differences in template DNA concentrations of barcoding regions (such as ITS) that fall within the rDNA cassette. Given our demonstration that rDNA copy number can vary widely among closely related taxa, comparisons based on ITS sequence read abundance even among members of the same species may strongly over- or underestimate actual individual fungal abundance. Efforts to account for these effects, as has been applied in other microorganisms (Kembel et al., 2012; Stoddard et al., 2015), remain a significant research priority.

In other eukaryotic organisms (e.g., plants and animals), rDNA CNV has been shown to have a strong positive correlation with genome size, independent of size contributions from rDNA cassettes (Prokopowich et al., 2003; Wencai et al., 2018). Conversely, investigation into this correlation for bacteria has shown no such relationship (Fogel, Collins, Li, & Brunk, 1999). Contrary to other eukaryotes, we found no indication that rDNA CNV is related to genome size in fungi (and regardless of whether or not rRNA cassette size was also considered). The finding that fungi do not conform to the pattern recognized between rDNA CNV and genome size may offer an interesting counterpoint for future analyses into the mechanisms structuring this relationship in plants and animals.

Using an in silico approach coupled with computational benchmarking, we have demonstrated that fungi exhibit substantial rDNA CNV that is inversely correlated with phylogenetic relatedness. While we did not observe strong links between rDNA CNV and ecological lifestyle, the continued use of this analysis pipeline on the rapidly increasing number of fungal genomes being generated will enable greater consideration of this trait in future studies. Similarly, using this pipeline in conjunction with studies characterizing rDNA gene expression will further enhance our understanding of fungal responses to a broad range of environmental conditions. Importantly, the range of rDNA copy numbers estimated for fungi, which have often been thought to bridge the macro- and microbiological worlds, falls between lower rDNA copy numbers in prokaryotes and higher rDNA copy numbers in many other eukaryotes. As such, identifying the evolutionary and ecological mechanisms constraining CNV for fungi will help facilitate a broader understanding of the influence of rDNA CNV across all domains of life.

ACKNOWLEDGEMENTS

We thank R. Vilgalys for sharing cultures and access to the unpublished genomic data of four fungal species, A. Gryganskyi for use of unpublished genomic data for three species of early diverging fungi, R. de Vries, M. Mäkelä and K. Hildén for use of unpublished genomic data for the *D. squalens* sequencing projects, I. Grigoriev and the JGI MycoCosm team for assistance accessing fungal genomes, F. Keck for assistance with using Phylosignal and 3 anonymous reviewers for helpful comments and feedback. Support for this study was provided by a National Science Foundation (NSF) Graduate Research Fellowship to L.A. Lofgren, NSF grant #1441604 to J. Uehling, and NSF grant #1554375 to P.G. Kennedy.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

P.G.K., L.A.L. and J.K.U. conceived the study, L.A.L. performed the computational and phylogenetic analyses, and S.B., T.D.B., P.G.K., L.A.L. and F.M. contributed to sequencing. P.G.K. and L.A.L. wrote the manuscript with feedback and editorial advice from J.K.U., T.D.B., S.B. and F.M.

DATA ACCESSIBILITY

All unpublished data included in this project were used with permission of the project PIs. Metadata for each genome sequencing project can be found at the JGI genome pages for each sequencing project. All code and data associated with this project have been made open access and can be found on the project's GitHub at https://github.com/MycoPunk/rDNA_CNV.

ORCID

Lotus A. Lofgren https://orcid.org/0000-0002-0632-102X
Jessie K. Uehling https://orcid.org/0000-0003-0439-2737
Sara Branco https://orcid.org/0000-0002-9890-933X
Thomas D. Bruns https://orcid.org/0000-0002-2943-8669
Francis Martin https://orcid.org/0000-0002-4737-3715
Peter G. Kennedy https://orcid.org/0000-0003-2615-3892

REFERENCES

Baldrian, P., Větrovský, T., Cajthaml, T., Dobiášová, P., Petránková, M., Šnajdr, J., & Eichlerová, I. (2012). Estimation of fungal biomass in forest litter and soil. *Fungal Ecology*, *6*, 1–11. https://doi.org/10.1111/nph.13849

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chiang, D. Y., Getz, G., Jaffe, D. B., O'Kelly, M. J. T., Zhao, X., Carter, S. L., … Lander, E. S. (2009). High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods*, *6*, 99–103. https://doi.org/10.1038/nmeth.1276

Edgar, Robert C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5.1: 113.

Elser, J. J., Sterner, R. W., Gorokhova, E., Fagan, W. F., Markow, T. A., Cotner, J. B., … Weider, L. W. (2000). Biological stoichiometry

from genes to ecosystems. *Ecology Letters*, *3*, 540–550. https://doi.org/10.1111/j.1461-0248.2000.00185.x

Fogel, G. B., Collins, C. R., Li, J., & Brunk, C. F. (1999). Prokaryotic genome size and SSU rDNA copy number: Estimation of microbial relative abundance from a mixed population. *Microbial Ecology*, *38*, 93–113. https://doi.org/10.1007%2Fs002489900162

Foster, I. (2006). Globus toolkit version 4: Software for service-oriented systems. *Journal of Computer Science and Technology*, *21*, 513–520. https://doi.org/10.1007%2Fs11390-006-0513-y

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*, 113–118. https://doi.org/10.1111%2Fj.1365-294x.1993.tb00005.x

Gong, J., Dong, J., Liu, X., & Massana, R. (2013). Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of Oligotrich and Peritrich ciliates. *Protist*, *164*, 369–379. https://doi.org/10.1016/j.protis.2012.11.006

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., … Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*, D699–704. https://doi.org/10.1093/nar/gkt1183

Henk, D. A., & Fisher, M. C. (2012). The gut fungus basidiobolus ranarum has a large genome and different copy numbers of putatively functionally redundant elongation factor genes. *PLoS ONE*, *7*, e31268. https://doi.org/10.1371/journal.pone.0031268

Ide, S., Miyazaki, T., Maki, H., & Kobayashi, T. (2010). Abundance of ribosomal RNA gene copies maintains genome integrity. *Science*, *327*, 693–696. https://doi.org/10.1126/science.1179044

Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, *6*(9), 2774–2780.

Kembel, S. W., Wu, M., Eisen, J. A., & Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, *8*, e1002743. https://doi.org/10.1371/journal.pcbi.1002743

Klappenbach, J. A., Dunbar, J. M., & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, *66*, 1328–1333. https://doi.org/10.1128%2Faem.66.4.1328-1333.2000

Kobayashi, T., Heck, D. J., Nomura, M., & Horiuchi, T. (1998). Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: Requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes and Development*, *12*, 3821–3830. https://doi.org/10.1101%2Fgad.12.24.3821

Kohler, A., Kuo, A., Nagy, L. G., Morin, E., Barry, K. W., Buscot, F., … Martin, F. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics*, *47*, 410–415. https://doi.org/10.1038/ng.3223

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao, D. (2000). Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea. *Journal of Molecular Evolution*, *51*, 305–317. https://doi.org/10.1007%2Fs002390010093

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., … Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*, 337–341. https://doi.org/10.1038/nature07743

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*, 884–890. https://doi.org/10.1038/ng.2678

Maleszka, R., & Clarkwalker, G. D. (1993). Yeasts have a 4-fold variation in ribosomal DNA copy number. *Yeast*, *9*, 53–58. https://doi.org/10.1002%2Fyea.320090107

Martino, E., Morin, E., Grelet, G. A., Kuo, A., Kohler, A., Daghino, S., … Perotto, S., et al. (2018). Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytologist*, *217*, 1213–1229. https://doi.org/10.1111/nph.14974

McKerracher, L., & Heath, I. (1985). The structure and cycle of the nucleus-associated organelle in two species of Basidiobolus. *Mycologia*, *77*, 412–417. https://doi.org/10.2307%2F3793197

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, Ieee, *2010*.

Miyazaki, T., & Kobayashi, T. (2011). Visualization of the dynamic behavior of ribosomal RNA gene repeats in living yeast cells. *Genes to Cells*, *16*, 491–502. https://doi.org/10.1111%2Fj.1365-2443.2011.01506.x

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*, 743–756. https://doi.org/10.1111/j.2041-210x.2012.00196.x

Nemergut, D. R., Knelman, J. E., Ferrenberg, S., Bilinski, T., Melbourne, B., Jiang, L., … Townsend, A. R. (2016). Decreases in average bacterial community rRNA operon copy number during succession. *ISME Journal*, *10*, 1147–1156. https://doi.org/10.1038%2Fismej.2015.191

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., & Schilling, J. S. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2014). Caper: comparative analyses of phylogenetics and evolution in R. R package version 0.5.2/ r121, 1–36.

Paredes, S., Branco, A. T., Hartl, D. L., Maggert, K. A., & Lemos, B. (2011). Ribosomal DNA deletions modulate genome-wide gene expression: rDNA-sensitive genes and natural variation. *PLoS Genetics*, *7*, e1001376. https://doi.org/10.1371/journal.pgen.1001376

Perisin, M., Vetter, M., Gilbert, J. A., & Bergelson, J. (2016). 16Stimator: Statistical estimation of ribosomal gene copy numbers from draft genome assemblies. *The ISME Journal*, *10*, 1020–1024. https://doi.org/10.1038/ismej.2015.161

Prokopowich, C. D., Gregory, T. R., & Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, *46*, 48–50. https://doi.org/10.1139%2Fg02-103

Pukkila, P. J., & Skrzynia, C. (1993). Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete Coprinus cinereus. *Genetics*, *133*, 203–211.

R Core Team (2017). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org/.

Rehner, S., & Samuels, G. (1995). Molecular systematics of the Hypocreales: A teleomorph gene phylogeny and the status of their anamorphs. *Canadian Journal of Botany*, *73*, 816–832. https://doi.org/10.1139/b95-327

Rogers, S. O., & Bendich, A. J. (1987). Heritability and variability in ribosomal RNA genes of Vicia faba. *Genetics*, *117*, 285–295.

Russell, P. J., & Rodland, K. D. (1986). Magnification of rRNA gene number in a Neurospora crassa strain with a partial deletion of the nucleolus organizer. *Chromosoma*, *93*, 337–340. https://doi.org/10.1007%2Fbf00327592

Salim, D., Bradford, W. D., Freeland, A., Cady, G., Wang, J., Pruitt, S. C., & Gerton, J. L. (2017). DNA replication stress restricts ribosomal DNA copy number. *PLoS Genetics*, *13*, e1007006. https://doi.org/10.1371/journal.pgen.1007006

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., … Probst, A. V. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in Arabidopsis thaliana. *Nucleic Acids Research*, *46*, 3019–3033. https://doi.org/10.1093%2Fnar%2Fgky163

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690. https://doi.org/10.1093%2Fbioinformatics%2Fbtl446

Stevenson, B. S., & Schmidt, T. M. 2004. Life history implications of rRNA gene copy number in *Escherichia coli*. *Society 70*: 6670–6677. https://doi.org/10.1128/aem.70.11.6670-6677.2004

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., & Schmidt, T. M. (2015). rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, *43*, D593–D598. https://doi.org/10.1093%2Fnar%2Fgku1201

Szostak, J. W., & Wu, R. (1980). Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature*, *284*, 426–430. https://doi.org/10.1038%2F284426a0

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., … Wilkens-Diehr, G. D. (2014). XSEDE: Accelerating scientific discovery. *Computing in Science and Engineering*, *16*, 62–74. https://doi.org/10.1109%2Fmcse.2014.80

Wencai, W., Tao, W., Aleš, K., Becher, H., Kuderova, A., Leitch, I., … Leitch, A. (2018). Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Annals of Botany*, *00*, myc172. https://doi.org/10.1093/aob/mcy172

White, T. J., Bruns, S., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phyologenetics.pdf. *PCR Protocols: A Guide to Methods and Applications*, 315–322. https://doi.org/10.1016%2Fb978-0-12-372180-8.50042-1

Yoon, S., Xuan, Z., Makarov, V., Ye, K., & Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research*, *19*, 1586–1592. https://doi.org/10.1101%2Fgr.092981.109

Zolan, M. E. (1995). Chromosome-length polymorphism in fungi. *Microbiological Reviews*, *59*, 686–698. https://doi.org/10.1016%2Fj.mimet.2006.05.022

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: 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. *Mol Ecol*. 2019;28:721–730. https://doi.org/10.1111/mec.14995