Trends in Microbiology

CellPress REVIEWS

Opinion

Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities?

Anton Lavrinienko,[1,3] Toni Jernfors,[1,3] Janne J. Koskimäki,[2] Anna Maria Pirttilä,[2] and Phillip C. Watts[1,*]

Amplicon sequencing of partial regions of the ribosomal RNA loci (rDNA) is widely used to profile microbial communities. However, the rDNA is dynamic and can exhibit substantial interspecific and intraspecific variation in copy number in prokaryotes and, especially, in microbial eukaryotes. As change in rDNA copy number is a common response to environmental change, rDNA copy number is not necessarily a property of a species. Variation in rDNA copy number, especially the capacity for large intraspecific changes driven by external cues, complicates analyses of rDNA amplicon sequence data. We highlight the need to (i) interpret amplicon sequence data in light of possible interspecific and intraspecific variation, and (ii) examine the potential plasticity in rDNA copy number as an important ecological factor to better understand how microbial communities are structured in heterogeneous environments.

Highlights

Amplicon sequencing of part of the ribosomal RNA locus (hereafter, rDNA) is a widespread methodology that has uncovered vast diversity and macroecological patterns in microbial communities.

While interspecific variation in rDNA copy number may complicate analyses of microbial communities using amplicon sequence data, the occurrence of intraspecific variation in rDNA copy number adds an extra dimension of complexity.

As intraspecific variation in rDNA copy number is associated with environment variation, apparent demographic changes in a microbial community may be driven by a genomic response to the environment.

Intraspecific variation in rDNA copy number may be a greater problem for studies of microbial eukaryotes than prokaryotes, and hence the challenge in interpreting amplicon sequence data.

Quantifying Microbial Communities Using Molecular Methods

Next-generation sequencing (**NGS**) (see Glossary) technology has provided the means to identify and count microbial taxa within samples at a spatiotemporal scale that is impractical using culture-based methods and/or Sanger sequencing, as exemplified by global assessments of eukaryote [1–3] and prokaryote [4,5] microbial communities. Use of NGS to quantify microbial communities, nonetheless, is accompanied by certain technical pitfalls – such as biases associated with library preparation, biases in the choice of NGS platform, and/or biases that occur during **PCR amplification** [6,7]. Many such technical issues have been addressed by the development of standard protocols (e.g., [4,5]) and bioinformatics pipelines [8]. Here, we highlight how interspecific and intraspecific variation in **ribosomal RNA** (**rRNA**) locus (Figure 1) copy number may confound analyses of microbial community composition, especially when the focus is on eukaryotic microbes whose genomes can exhibit extensive interspecific and intraspecific variation in rRNA locus copy number.

Part of the Ribosomal RNA Gene Cluster (rDNA) Is the Amplicon of Choice to Quantify Microbial Community Composition

Typical NGS analyses of microbial community composition use **amplicon sequencing** (or **marker gene analysis**) [8], where the end products of resolved sequences (**operational taxonomic units** (**OTUs**) or **exact sequence variants** (**ESVs**) [9]) are compared with a DNA **reference library** to assign their taxonomic identities [8]. The appropriate region(s) of the genome from which amplicons are derived depends on the level of interspecific and intraspecific sequence divergence and availability of **'universal' PCR primers**. Amplicons typically are derived from part of the cluster of rRNA loci (**rDNA**) (Figure 1), such as one of the variable regions of the 16S rDNA for prokaryotes [6], a variable region of the 18S or 28S rDNA [10,11] or in one of

[1]Department of Biological and Environmental Science, University of Jyväskylä, 40014 Jyväskylä, Finland
[2]Ecology and Genetics, University of Oulu, 90014 Oulu, Finland
[3]These authors contributed equally to this work

*Correspondence:
phillip.c.watts@jyu.fi (P.C. Watts).

the rDNA's **internal transcribed spacers** (**ITSs**) [12,13] when quantifying eukaryotic microbial community diversity. After assigning taxonomy to ESVs, the next fundamental step in an analysis of amplicon sequence data is to count the number of sequences (NGS reads) belonging to each ESV: here, a key assumption is that the proportion of reads assigned to each ESV reflects the relative abundance (e.g., number of cells or biomass) of putative taxa within the sample. However, this assumption is complicated by the fact that rDNA is organized as a **tandem array** in many species (Figure 1) and this region of the genome can exhibit substantial interspecific and intraspecific variation in copy number.

There is a marked contrast in the level of interspecific variation in rDNA copy number among taxonomic Domains. Prokaryotes typically have fewer than seven rDNA copies (median = 5 and 1 rDNA copies for Bacteria (n = 15 486 genomes) and Archaea (n = 343 genomes) respectively, (*rrn*DB v.5.6 [14], date accessed 22 May 2020), albeit with one strain of the bacterium *Photobacterium damselae* having as many as 21 copies of 16S rDNA. In contrast, eukaryotic rDNA copy number exhibits extensive interspecific variation. For example, rDNA copy number is estimated to vary from 14 to 1442 copies in fungi [15] and between 1 and >500 000 copies amongst species of protist [16–20], with notably high estimates of rDNA copy number per cell for ciliates [16–18]. rDNA copy number is positively correlated with eukaryotic genome size [21], although this association may not hold for ciliates and fungi [15,16]; other studies found a positive association between rDNA content and cell size in some marine protist taxa [18,19]. Why rRNA locus exhibits such interspecific diversity, and is often one of the most abundant regions of the eukaryotic genome, is a complex issue, related to regulation of rRNA transcription, nucleolus function, and other cellular processes [22–25]. From a community ecology perspective, however, extensive interspecific variation in rDNA copy number limits the efficacy of NGS-based methods to accurately enumerate the relative proportions of microbial taxa within a sample [26,27].

In theory, better estimates of taxonomic proportions from rDNA amplicon data may be recovered by adjusting the counts of ESVs with taxon-specific values of rDNA copy number per genome, and some software can implement this procedure for prokaryote samples (e.g., [26,27]). In practice, rDNA copy number is unknown for most taxa and this type of bioinformatic correction relies on an apparent phylogenetic conservation of rDNA copy number [26,28] which, in prokaryotes, may exist over short phylogenetic distances only [27,29]. Similarly, rDNA content was often comparable among congeneric fungi, but with frequent exceptions [15]. For microbial eukaryote taxa, attempting to correct counts of ESVs is not feasible as the rDNA copy number in genomes of sufficient species and the extent of any phylogenetic conservation of rDNA copy number is unknown. The impact of interspecific variation in rDNA copy number remains an unresolved issue in the analysis of microbial species' proportions from amplicon data [27], especially for analyses of eukaryotic communities. Even with taxon-specific data on rDNA copy number per genome, however, molecular analyses of microbial community composition may be compromised by the occurrence of intraspecific variation in rDNA copy number.

Intraspecific Variation in rDNA Copy Number in Prokaryotes and Eukaryotes

Given its essential functions, rDNA copy number is tightly regulated [30]. And yet, the rRNA locus represents a notably dynamic region of the genome [22–25] that exhibits wide intraspecific variation in copy number. Intraspecific variation in rDNA copy number is not widely reported in studies of prokaryotes [14], although some bacteria tolerate a change in rDNA copy number [31–33] such as an expansion to 17 rDNA copies in the genome of *Paeniclostridium sordellii* CBA7122, whose genome usually contains an average of four 16S rDNA copies [34]. In

Glossary

Amplicon sequencing: deciphering the sequence of amplified DNA fragments.
Concerted evolution: a process by which multiple related loci are homogenized (or evolved in concert) so that the DNA sequences within a species shares higher identity than when compared between species.
Droplet digital PCR (ddPCR): a method that provides the ability to quantify the amounts of nucleic acids (digital PCR) in an individual droplet out of 20 000 emulsified droplets.
Exact sequence variant (ESV): signifies the use of the exact DNA sequences originating from the reads, rather than clustering reads into operational taxonomic units (OTUs).
Internal transcribed spacer region (ITS): the untranslated DNA sequence between ribosomal RNA genes.
Marker gene analysis: sequencing of short fragments of DNA from a gene or genes that exhibit significant sequence variability and divergence to be used for species identification.
Metagenome-assembled genome (MAG): a single-taxon assembly based on computational binning (or classification) of contiguous sequences with similar properties.
Next-generation sequencing (NGS): fast and efficient method to perform sequencing of millions of fragments of DNA in a massively parallel reaction.
Operational taxonomic unit (OTU): a cluster of sequences of a specific taxonomic marker gene grouped according to their similarity (typically, the similarity threshold is 97%).
PCR amplification: the polymerase chain reaction is a method for exponentially amplifying copies of a particular DNA segment.
rDNA: a segment of genome consisting of the loci for ribosomal RNAs and spacer sequences (often arranged in tandem repeats).
Reference library: an annotated collection of DNA sequences that can be used to resolve sequence identity in the NGS-generated data.
Ribosomal RNA (rRNA): comprises the RNA molecules that are structural components of ribosomes, and which are encoded by ribosomal DNA genes.
Single-amplified genome (SAG): a single-taxon assembly generated with single-cell sequencing; it requires physical isolation of individual cells, their whole-genome amplification, and subsequent sequencing.

eukaryotes, by contrast, substantial intraspecific variation in rDNA copy number appears commonplace [24,25]. In a survey of 4876 strains of baker's yeast (*Saccharomyces cerevisiae*), rDNA copy number varied from less than 80 to more than 450 copies among mutants [35,36]; isolates of other fungal species have uncovered a twofold to fourfold variation in rDNA copy number [37,38]. Ciliates maintained in laboratory cultures also exhibit marked changes in rDNA copy number [17,39,40], for example between an estimated 1082 and 16 995 copies (~15-fold) in *Strombidium stylifer* [16]. Hence, there is apparently greater potential for microbial eukaryotes to exhibit substantial intraspecific variation in rDNA copy number compared with prokaryotes.

Tandem array: gene copies arranged in tandem repeats in a genome generated by tandem duplications.
Universal PCR primers: short pieces of DNA (primers) that can be used to simultaneously amplify DNA of diverse taxa using PCR.

A further consideration in rDNA copy number variation is the occurrence of intragenomic polymorphism. Despite the potential for **concerted evolution** to reduce rDNA sequence divergence [25], intragenomic rDNA polymorphisms have been reported, for example, in almost 50% of examined bacteria and some 3–5% of fungi [41,42]. The distribution of any rDNA polymorphisms among multiple rDNA copies within genomes of many taxa is largely unknown. Use of OTU-based clustering (rather than ESVs) to define taxa could minimize the potential impact of intragenomic rDNA variation on analysis of microbial communities using amplicon sequence data [42–45].

rDNA Is Sensitive to Environmental Variation

Understanding the function of rDNA copy number variation is an important part of interpreting patterns of community structure derived from rDNA amplicon sequence data. The adaptive significance of variation in rDNA copy number is well studied in bacteria, for example, being associated with interspecific differences in metabolism [24,46] and growth rate [32,47,48], and also acting as a trait associated with habitat specialization [48] or community succession [49,50]. Few studies have examined the potential significance of interspecific variation in rDNA copy number in natural eukaryotic microbial communities, even though rDNA copy number predicts sensitivity to DNA damage [23,51] and may explain species' response to stress in laboratory settings [22,52].

What is particularly relevant for analyses of microbial community composition is that rDNA copy number is not necessarily a species-level trait, given widespread evidence that certain stimuli can elicit rapid intraspecific changes in rDNA copy number in microorganisms (Table 1). Conversely, rDNA copy number in isolates of *Aspergillus fumigatus* was stable when exposed to a fungicide or changes in temperature [38]; the implication is that not every species undergoes a rapid and/or detectable plasticity in rDNA copy number. Indeed, the lack of taxonomic diversity in Table 1 highlights the need to better quantify the extent to which rDNA copy number is a species' trait or varies in response to environmental cues. Moreover, there is a need to better document the types of environmental cue that elicit a general or taxon-specific change in rDNA copy number. While it is challenging to correct amplicon sequence data for interspecific variation in rDNA copy number, that the environment per se can impact rDNA copy number adds another layer of complexity (Figure 2, Key Figure) that is typically not considered in surveys of natural microbial eukaryote communities. Whether any intragenomic rDNA genotypes alter their copy number differentially in response to environmental variation is not known, but whose impact on any analysis of amplicon sequence data depends on the level of sequence divergence among rDNA polymorphisms and whether taxa are defined as OTUs or ESVs. Analyses of rDNA amplicon data, especially in eukaryotic microbes, should consider the likelihood of an intraspecific genomic response to the environment and its potential interaction with a species' rDNA copy number. This is an important consideration given that the aim of many studies is to quantify changes in community composition in response to a change in environment that itself may stimulate a change in rDNA architecture.

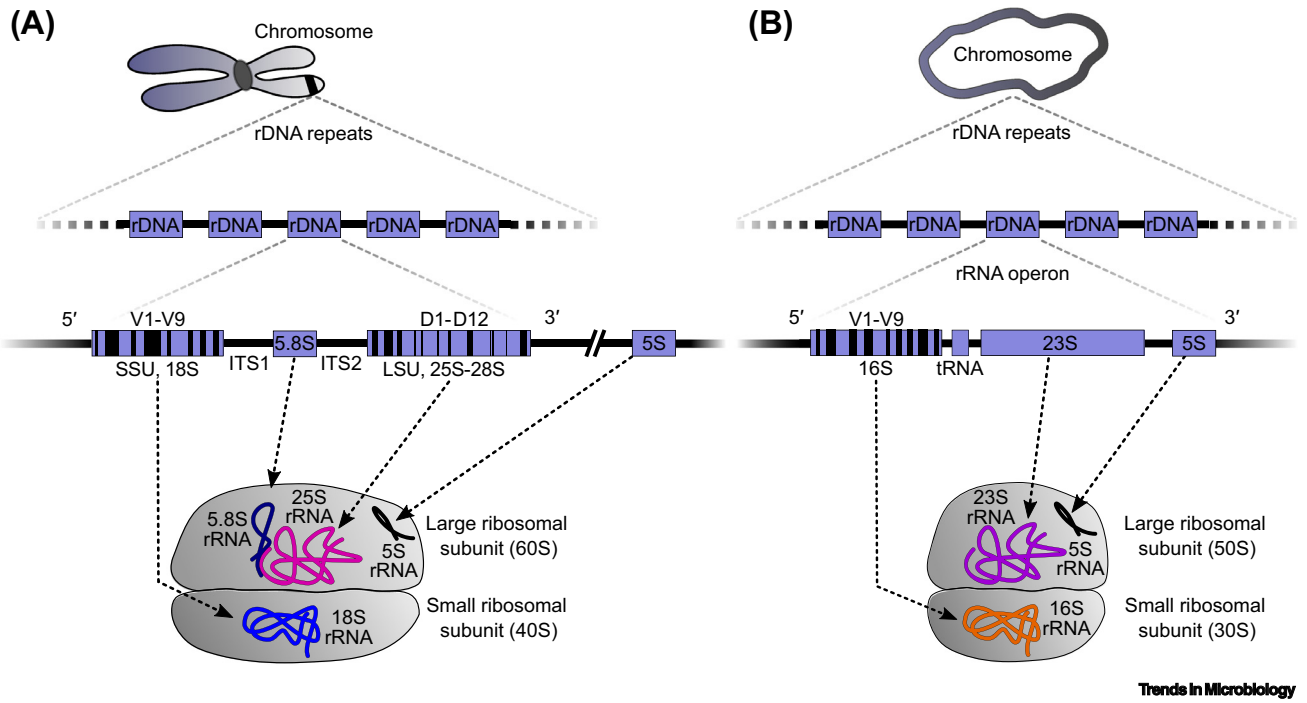


Figure 1. Schematic Representation of the Ribosomal RNA (rRNA) Gene Cluster (or rDNA). The variable regions of (A) eukaryotic and (B) prokaryotic rRNA loci are commonly used to characterize microbial taxa and resolve their phylogenetic relationships. In most fungi, the rRNA gene cluster includes the small ribosomal subunit (SSU, 18S), with internal transcribed spacer regions (ITS1 and ITS2) flanking the 5.8S, and large ribosomal subunit (LSU, 25–28S) regions. In bacteria, the rRNA operon comprises the SSU (16S), LSU (23S), and 5S loci. Black vertical lines in serial order illustrate the variable regions in SSU (V1–V9) and LSU (D1–D12), best suited for biodiversity assessments through microbial communities profiling.

Does rDNA Amplicon Sequencing Have a Future for Assessments of Microbial Community Composition?

rDNA is an excellent target for amplicon sequencing because the genomes of all living organisms have homologous loci, and investment into the design of universal PCR primers has enabled the use of a single methodology to identify diverse taxa. An important corollary of the historical use of rDNA sequence data to resolve phylogenetic relationships and identify taxa is the many, large and curated reference databases of rDNA sequences [53–56] that provide a standardized method of assigning taxonomy to microbial ESVs/OTUs. The potential for environment-driven changes in rDNA copy number within species does not make this locus redundant for assessments of community composition but it highlights the need to consider more deeply what the community responses to the environment might be: demographic, genomic, or a combination of the two (Figure 2).

Current solutions to obtain better taxonomic proportions using rDNA amplicon sequence data emphasize a need for more data on rDNA copy number in more species. Interspecific differences in rDNA copy number can be examined in prokaryotes using the *rrn*DB database [14]. Levels of intraspecific variation in rDNA copy number are not addressed in *rrn*DB, but could be evaluated for key species (e.g., of medical importance) by mapping NGS read data to assembled genomes and/or with long-read sequencing technology to better assemble rDNA operons [57]. Such effort may not be warranted as a general strategy to improve analyses of prokaryote community composition, however, given the comparatively low level of interspecific and intraspecific rDNA copy number variation among prokaryotes (where about 50% of 15 829 records in *rrn*DB (v.5.6, date accessed 22 May 2020) [14] have four or fewer rDNA copies per genome).

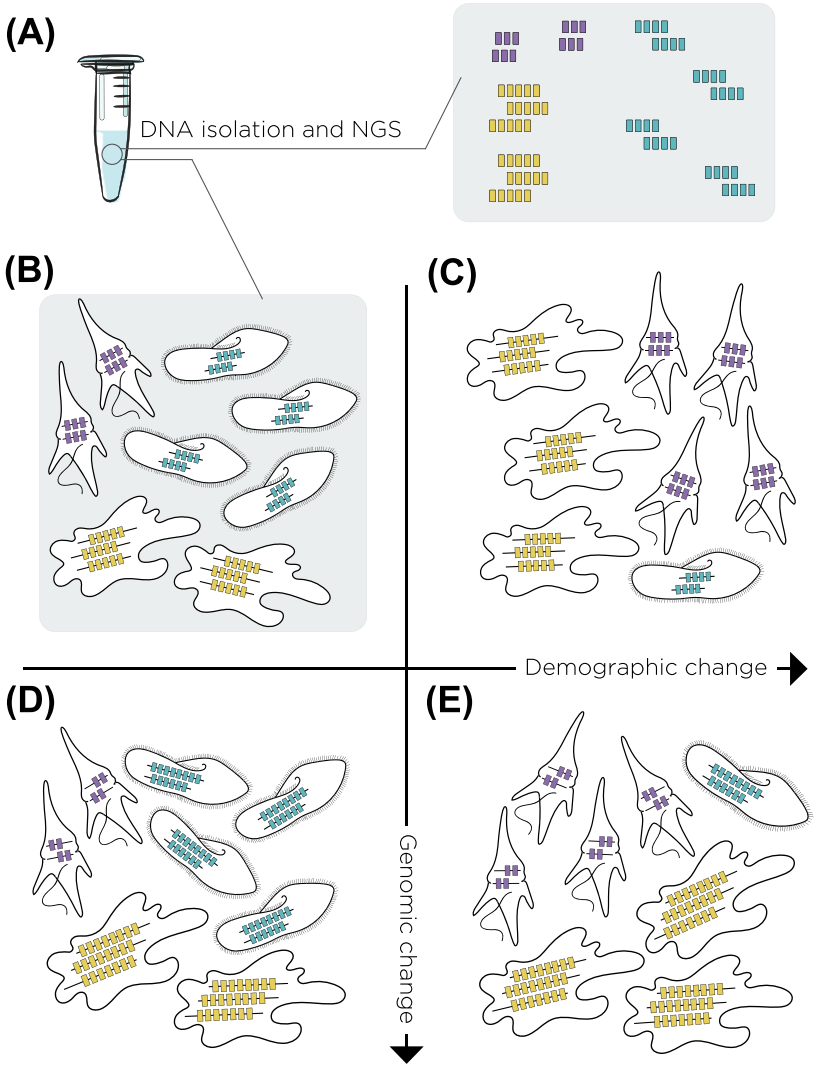
Table 1. Studies Examining Intraspecific Variation in rDNA Copy Number (CN) and Fitness Correlations in Prokaryotes and in Microbial Eukaryotes

Species	Source of variation	Observation	Refs
Prokaryotes			
Bacillus subtilis	Natural variation	rDNA CN variation observed between *Bacillus subtilis* strains	[33]
Escherichia coli	Induced deletion	Deletion of rDNA copies results in lower fitness in competition cultures in resource rich medium	[32]
Eukaryotes			
Budding yeast	Chemical stressors	Strains with lower rDNA CN than normal are more sensitive to DNA damaging agents	[51]
Budding yeast	Resource richness	Activation of TOR pathway as a result of nutrient rich medium increases rate of rDNA amplification	[68]
Budding yeast	Chemical stressors	Induced replication stress selects for lower rDNA CN	[52]
Mold *Aspergillus fumigatus*	Temperature and chemical stressors	*Aspergillus* mold showed stable rDNA CN across various treatments such as temperature and exposure to antifungal agents	[38]
Ciliate *Tetrahymena puriformis*	Resource richness	Cells in exponential growth phase contain an increased amount of rDNA compared with starved cells or cells in the stationary growth phase	[69]
Ciliates *Entodinium, Epidinium* and *Ophryoscolex*	Resource richness	rDNA CN per cell changes in response to nutrient availability in starved and fed cultures	[70]
Ciliate *Chilodonella uncinata*	Artificial evolution, bottlenecking	rDNA CN does not follow assumptions of stochasticity in artificial evolution experiment, exhibiting directional selection with increased CN	[39]
Ciliates *Halteria grandinella, Strombidium stylifer* and *Blepharisma americanum*	Natural variation	Extreme interspecific and intraspecific variation of rDNA CN observed in various ciliate species	[16]
Ciliates *Euplotes vannus* and *Strombidium sulcatum*	Temperature	Negative association between rDNA content and temperature between 16°C and 25°C, with a loss of about 6000–8000 (ca. 3.5%) rDNA copies/°C	[18]

Given extensive variation in rDNA copy number, quantifying taxonomic composition of eukaryotic microbial communities using rDNA amplicon sequence data is challenging. Prospects of developing a bioinformatic solution to account for variation in rDNA copy number in eukaryotes appear poor. While an rDNA copy number database exists for animals [58], rDNA copy number data are lacking for microbial eukaryotes. Developing an rDNA copy number resource for microbial eukaryotes is difficult given (i) the diversity of protist species [59], (ii) difficulties in isolating and/or culturing many species, (iii) our ignorance of the environmental cues that drive changes in rDNA copy number, and/or (iv) the likelihood that there is no phylogenetic conservation of rDNA copy number variation [15,28]. Use of the apparent cell size-rDNA locus copy number relationship [19] to adjust read counts to better reflect abundance of certain protist taxa will likely introduce more noise into the analysis given (i) the inherent variance in this relationship and (ii) the potential for intraspecific variation in rDNA content. An alternative method of resolving community composition in environmental DNA samples would be to use single-copy loci [60,61] as targets for amplicon sequencing, although it would take a substantial effort to identify a

Key Figure

Possible Scenarios That Can Alter the Ribosomal RNA Loci (or rDNA) Composition of a Microbial Community



	Scenario							
	B		C		D		E	
Taxa	*n*	*N*rDNA	*n*	*N*rDNA	*n*	*N*rDNA	*n*	*N*rDNA
taxon1	2	12	4	24	2	8	4	16
taxon2	2	30	3	45	2	46	3	69
taxon3	4	32	1	8	4	56	1	14

Trends in Microbiology

(See figure legend at the bottom of the next page.)

panel of reliable loci that have (i) an appropriate level of sequence divergence for taxonomic assignment and (ii) sufficiently conserved regions that allow the design of universal primers that yield suitable amplicon lengths. Abundances of key strains may be quantified using quantitative PCR (qPCR) or **droplet digital PCR** (ddPCR) [62], but these methods are impractical for analyses of whole communities. Metagenomic sequencing offers an alternative to analysis of rDNA amplicons as reads can be mapped to metagenomes whose taxonomic identities are ascertained using a panel of conserved loci [61,63]. Metagenomics data could be a useful source of rDNA copy number information from organisms derived directly from environmental samples bypassing cultivation. That said, difficulties in assembly and binning of repetitive genome regions (such as rDNA) limit the use of **metagenome-assembled genomes** (**MAGs**) or **single-amplified genomes** (**SAGs**) for quantifying rDNA copy number variation [64]. Long-read sequencing technologies can overcome this issue by improving assembly contiguity or even completing the genome [65]. A hybrid approach that combines short- and long-read sequencing will likely yield more complete genomes from metagenomes in future studies, thus populating databases with rDNA copy number from microorganisms in natural systems [64]. However, even with the rapid advances in NGS technology and bioinformatics pipelines, metagenomic sequencing is currently too resource intensive (e.g., time consuming and expensive, see [66]) to use as a method to profile community composition or quantify rDNA content in many samples. Thus, studies that use the rDNA-based amplicon sequencing to quantify a change in community profile need to interpret their data in light of possible interspecific and intraspecific responses.

Concluding Remarks

A corollary of the long-term investment in developing standard laboratory protocols and large, curated rDNA databases (described earlier) is that rDNA amplicon sequencing remains a straightforward and cost-effective method of quantifying microbial community composition. rDNA structure and dynamics is less well studied in the genomes of microbial eukaryotes in natural systems compared with prokaryotes, and there appears to be potential for greater interspecific and intraspecific variation in rDNA copy number in eukaryote genomes. This variation in rDNA copy number, especially the capacity for large intraspecific changes, complicates analysis of rDNA amplicon sequence data, but instead of being a nuisance parameter could be viewed as a positive challenge to make the most of trait-based approaches in microbial ecology [67] (see Outstanding Questions). In prokaryotes, for example, analyses of rDNA copy number often extend beyond its use as a barcode locus, to examine rDNA architecture as an important trait related to ecological strategy [24,32,46–50]. Analyses of natural eukaryote microbial communities would likewise benefit from identifying the responses of taxa, whether they change their relative abundance and/or alter their rDNA copy number in response to changes in the environment (Figure 2). Isolation and culturing of such taxa would allow the use of laboratory experiments and qPCR analyses to examine the role of rDNA dynamics in eukaryotic community dynamics. Integrating these approaches

Figure 2. (A) In microbial ecology, a typical next-generation sequencing (NGS) analysis starts from DNA extraction and sequencing, and results in a catalogue of exact sequence variants (ESVs) or operational taxonomic units (OTUs) that belong to different taxa, and which can be taxonomically resolved based on the rDNA sequence identity. (B) The overall rDNA content of a microbial community is a function of (i) the taxonomic composition (represented by taxon1, taxon2 and taxon3), and (ii) the rDNA copy number per genome (colored boxes within each taxon). A change in the environment may (C) alter the relative proportions of taxa (a demographic effect) or (D) elicit changes in the rDNA copy number per genome of each taxon (a genomic effect), or (E) affect both taxa proportions and their rDNA copy numbers. A comparison of the numbers of taxa (n) and the number of rDNA copies (N_{rDNA}) present in each scenario illustrates the potential difficulties associated with inferring relative proportions of taxa using ESV/OTU count data alone when there is interspecific and intraspecific variation in rDNA copy number.

Outstanding Questions

What is the capacity for intraspecific variation in ribosomal RNA locus (hereafter, rDNA) copy number in prokaryotes, particularly in nonlaboratory model taxa that can be isolated from natural systems?

Is there robust evidence that intraspecific variation in rDNA copy number is prevalent in microbial eukaryotes?

Is there a reliable predictor of the level of intraspecific variation in rDNA copy number, such as species' life history, ecological traits, phylogenic relationships, or genome size?

Is there a threshold value at which intraspecific variation in rDNA copy number starts to seriously affect rDNA-based analysis of microbial communities?

What are the most important environmental drivers of intraspecific variation in rDNA copy number?

What types of external stimuli affect rDNA copy number in many taxa, and what stimuli are species-specific?

What taxa are more (or less) susceptible to changes in rDNA copy number?

How does rDNA copy number and/or the capacity for rDNA copy number change mediate community structure, for example by affecting outcomes of competition or invasion success?

Are all components of the rRNA operon amplified in the same way when there are changes in copy number? For example, do nontranscribed regions, such as the ITS, have greater capacity for copy number variation than transcribed regions such as the 18S or 28S rRNA loci?

Does rDNA copy number have a predictable response (e.g., percent increase or decrease) to environmental variation?

with NGS surveys would provide key insights into understanding how microbial communities respond to variation and changes in the environment.

Acknowledgments

We are grateful for comments made by anonymous referees. This work was supported by the Academy of Finland (to P.C.W. projects 287153 and 329334; to A.M.P. and J.J.K. project 308766), University of Oulu Graduate School (A.L.), and Finnish Cultural Foundation (T.J.).

References

1. Tedersoo, L. *et al.* (2014) Global diversity and geography of soil fungi. *Science* 346, 1256688
2. Giner, C.R. *et al.* (2020) Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *ISME J.* 14, 437–449
3. Oliverio, A.M. *et al.* (2020) The global-scale distributions of soil protists and their contributions to belowground systems. *Sci. Adv.* 6, eaax8787
4. Sunagawa, S. *et al.* (2015) Structure and function of the global ocean microbiome. *Science* 348, 1261359
5. Thompson, L.R. *et al.* (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463
6. Gohl, D.M. *et al.* (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* 34, 942–949
7. Pollock, J. *et al.* (2018) The madness of microbiome: attempting to find consensus 'best practice' for 16S microbiome studies. *Appl. Environ. Microbiol.* 84, e02627-17
8. Knight, R. *et al.* (2018) Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422
9. Callahan, B.J. *et al.* (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643
10. Hadziavdic, K. *et al.* (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One* 9, e87624
11. Pawlowski, J. *et al.* (2012) CBOL Protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10, e1001419
12. De Filippis, F. *et al.* (2017) Different amplicon targets for sequencing-based studies of fungal diversity. *Appl. Environ. Microbiol.* 83, e00905-17
13. Schoch, C.L. *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6241–6246
14. Stoddard, S.F. *et al.* (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, 593–598
15. Lofgren, L.A. *et al.* (2019) Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Mol. Ecol.* 28, 721–730
16. Wang, C. *et al.* (2017) Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: impact of copy number variation and experimental error. *Proc. R. Soc. B Biol. Sci.* 284, 20170425
17. Gong, J. *et al.* (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
18. Zhu, F. *et al.* (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* 52, 79–92
19. Godhe, A. *et al.* (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl. Environ. Microbiol.* 74, 7174–7182
20. Medinger, R. *et al.* (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol. Ecol.* 19, 32–40
21. Prokopowich, C.D. *et al.* (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46, 48–50
22. Kobayashi, T. (2011) Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell. Mol. Life Sci.* 68, 1395–1403
23. Kobayashi, T. (2014) Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc. Japan Acad. Ser. B* 90, 119–129
24. Weider, L.J. *et al.* (2005) The functional significance of ribosomal (r)DNA variation: impacts on the evolutionary ecology of organisms. *Annu. Rev. Ecol. Evol. Syst.* 36, 219–242
25. Symonová, R. (2019) Integrative rDNAomics – importance of the oldest repetitive fraction of the eukaryote genome. *Genes (Basel)* 10, 345
26. Kembel, S.W. *et al.* (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput. Biol.* 8, e1002743
27. Louca, S. *et al.* (2018) Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* 6, 41
28. Angly, F.E. *et al.* (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2, 11
29. Langille, M.G.I. *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821
30. Nelson, J.O. *et al.* (2019) Mechanisms of rDNA copy number maintenance. *Trends Genet.* 35, 734–742
31. Sadeghifard, N. *et al.* (2006) The mosaic nature of intergenic 16S-23S rRNA spacer regions suggests rRNA operon copy number variation in *Clostridium difficile* strains. *Appl. Environ. Microbiol.* 72, 7311–7323
32. Stevenson, B.S. and Schmidt, T.M. (2004) life history implications of rRNA gene copy number in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 6670–6677
33. Widom, R.L. *et al.* (1988) Instability of rRNA operons in *Bacillus subtilis*. *J. Bacteriol.* 170, 605–610
34. Kim, J.Y. *et al.* (2017) Genomic analysis of a pathogenic bacterium, *Paeniclostridium sordellii* CBA7122 containing the highest number of rRNA operons, isolated from a human stool sample. *Front. Pharmacol.* 8, 840
35. Saka, K. *et al.* (2016) More than 10% of yeast genes are related to genome stability and influence cellular senescence via rDNA maintenance. *Nucleic Acids Res.* 44, 4211–4221
36. Kobayashi, T. and Sasaki, M. (2017) Ribosomal DNA stability is supported by many 'buffer genes' – introduction to the Yeast rDNA Stability Database. *FEMS Yeast Res.* 17, 1
37. Corradi, N. *et al.* (2007) Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Appl. Environ. Microbiol.* 73, 366–369
38. Herrera, M.L. *et al.* (2009) Strain-dependent variation in 18S ribosomal DNA copy numbers in *Aspergillus fumigatus*. *J. Clin. Microbiol.* 47, 1325–1332
39. Spring, K.J. *et al.* (2013) Chromosome copy number variation and control in the ciliate *Chilodonella uncinata*. *PLoS One* 8, e56413
40. Wang, Y. *et al.* (2019) Further analyses of variation of ribosome DNA copy number and polymorphism in ciliates provide insights relevant to studies of both molecular ecology and phylogeny. *Sci. China Life Sci.* 62, 203–214
41. Sun, D.L. *et al.* (2013) Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* 79, 5962–5969

42. Lindner, D.L. *et al.* (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecol. Evol.* 3, 1751–1764
43. Johnson, J.S. *et al.* (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10, 5029
44. Větrovský, T. and Baldrian, P. (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8, e57923
45. Nilsson, R.H. *et al.* (2019) Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat. Rev. Microbiol.* 17, 95–109
46. Roller, B.R.K. *et al.* (2016) Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* 1, 16160
47. Klappenbach, J.A. *et al.* (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333
48. Merhej, V. *et al.* (2009) Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4, 13
49. Nemergut, D.R. *et al.* (2016) Decreases in average bacterial community rRNA operon copy number during succession. *ISME J.* 10, 1147–1156
50. Shrestha, P.M. *et al.* (2007) Phylogenetic identity, growth-response time and rRNA operon copy number of soil bacteria indicate different stages of community succession. *Environ. Microbiol.* 9, 2464–2474
51. Ide, S. *et al.* (2010) Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* 327, 693–696
52. Salim, D. *et al.* (2017) DNA replication stress restricts ribosomal DNA copy number. *PLoS Genet.* 13, e1007006
53. del Campo, J. *et al.* (2018) EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol.* 16, e2005849
54. DeSantis, T.Z. *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072
55. Quast, C. *et al.* (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596
56. Nilsson, R.H. *et al.* (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, 259–264
57. Cuscó, A. *et al.* (2019) Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon. *F1000Research* 7, 1755
58. Sochorová, J. *et al.* (2018) Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. *Chromosoma* 127, 141–150
59. Larsen, B.B. *et al.* (2017) Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *Q. Rev. Biol.* 92, 229–265
60. Waterhouse, R.M. *et al.* (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548
61. Parks, D.H. *et al.* (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055
62. Morton, E.A. *et al.* (2020) Challenges and approaches to genotyping repetitive DNA. *G3 Genes Genomes Genet.* 10, 417–430
63. Tully, B.J. *et al.* (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* 5, 170203
64. Chen, L.X. *et al.* (2020) Accurate and complete genomes from metagenomes. *Genome Res.* 30, 315–333
65. Moss, E.L. *et al.* (2020) Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat. Biotechnol.* 38, 701–707
66. Bertrand, D. *et al.* (2019) Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* 37, 937–944
67. Lajoie, G. and Kembel, S.W. (2019) Making the most of trait-based approaches for microbial ecology. *Trends Microbiol.* 27, 814–823
68. Jack, C.V. *et al.* (2015) Regulation of ribosomal DNA amplification by the TOR pathway. *Proc. Natl. Acad. Sci. U. S. A.* 112, 9674–9679
69. Engberg, J. and Pearlman, R.E. (1972) The amount of ribosomal RNA genes in *Tetrahymena pyriformis* in different physiological states. *Eur. J. Biochem.* 26, 393–400
70. Sylvester, J.T. *et al.* (2009) Rumen ciliated protozoa decrease generation time and adjust 18S ribosomal DNA copies to adapt to decreased transfer interval, starvation, and monensin. *J. Dairy Sci.* 92, 256–269