Received: 16 July 2019 | Revised: 9 February 2020 | Accepted: 12 February 2020

DOI: 10.1002/edn3.77

ORIGINAL ARTICLE

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

The impact of intragenomic rRNA variation on metabarcoding-derived diversity estimates: A case study from marine nematodes

Tiago José Pereira[1] | Alejandro De Santiago[1] | Taruna Schuelke[1] | Sarah M. Hardy[2] | Holly M. Bik[1]

[1]Department of Nematology, University of California, Riverside, Riverside, CA, USA

[2]School of Fisheries and Ocean Sciences, University of Alaska, Fairbanks, AK, USA

Correspondence
Holly M. Bik, Department of Marine Sciences, University of Georgia, Athens, GA, USA.
Email: hbik@uga.edu

Present address
Tiago José Pereira, Alejandro De Santiago and Holly M. Bik, Department of Marine Sciences, University of Georgia, Athens, GA, USA
Taruna Schuelke, Department of Ecology, Evolution & Marine Biology, University of California Santa Barbara, Santa Barbara, CA, USA

Funding information
Gulf of Mexico Research Initiative; North Pacific Research Board, Grant/Award Number: 1303

Abstract

Intragenomic rRNA variation is a critical concern for eukaryotic metabarcoding studies, due to its potentially confounding effects on species delimitation and biodiversity estimates derived from -Omics data. In the present study, we assessed patterns associated with 18S rRNA metabarcoding loci in marine nematodes, including characterization of intragenomic rRNA gene variants (number of variants and abundance profiles) and aspects of datasets that can obscure biological signals (e.g., amplification of nontarget DNA, ambiguous taxonomy assignments). We estimated amplicon sequence variants (ASVs) using DADA2 from an 18S rRNA metabarcoding dataset (Illumina MiSeq) generated from individual marine nematodes. Illumina data were analyzed in conjunction with nematode morphological identifications and nearly full-length 18S reference sequences (~1,600 bp Sanger barcodes) generated for a subset of the same specimens. Our results indicated that levels of intragenomic rRNA variation appeared to vary widely across nematode taxa (irrespective of phylogenetic clades or ecological feeding groups) and that coamplification of nontarget DNA was common (relic DNA, gut contents, etc.). The DADA2 pipeline appeared to produce a biologically accurate profile of intragenomic rRNA variants in nematodes that was consistent with "Head-Tail" patterns (of dominant vs. minor rRNA gene variants) identified in previous studies. Although intragenomic rRNA variation appears to be ubiquitous in marine nematodes, nematode identifications were highly congruent across our three methods for species delimitation (traditional morphological taxonomy, Sanger DNA barcoding, and high-throughput metabarcoding). In spite of pervasive intragenomic variation and high copy number of rRNA genes, the most abundant ASVs in metabarcoding datasets are likely to represent true species barcodes and thus confer an accurate view of extant biodiversity. However, our findings also emphasize the importance of applying bioinformatic filtering techniques and developing well-curated reference databases in order to better link rRNA molecules

with specimen-level data and alleviate the confounding effects of intragenomic gene variants in studies of microbial eukaryotes.

KEYWORDS

18S rRNA gene, high-throughput sequencing, intragenomic variation, marine nematodes, metabarcoding

1 | INTRODUCTION

Advancements in high-throughput sequencing (HTS) technologies have drastically changed our way of exploring and characterizing the biodiversity of microscopic eukaryotes in freshwater, terrestrial, and marine ecosystems (Bik, Porazinska, et al., 2012; Bik, Sung, et al., 2012; Creer et al., 2010; Deiner et al., 2017; Fonseca et al., 2010; Geisen et al., 2018; Gibson et al., 2014; Orgiazzi, Dunbar, Panagos, de Groot, & Lemanceau, 2015). The rapidly decreasing per-sample cost, coupled with a corresponding increase in the volume of data generated from the newest platforms, has slowly begun to shift the taxonomic identifications of microbial eukaryotes from traditional morphology-based approaches toward environmental-Omics methods (Lallias et al., 2015; Orgiazzi et al., 2015; Pawlowski, Lejzerowicz, & Esling, 2014). Currently, HTS methods are being used for species discovery, comparisons of alpha- and beta-diversity patterns, and routine "eDNA biomonitoring" of communities and ecosystems (Baird & Hajibabaei, 2012; Deiner et al., 2017; Handley, 2015; Shaw et al., 2016; Thomsen & Willerslev, 2015).

Metabarcoding studies of microbial metazoa have overwhelmingly focused on ribosomal RNA (rRNA) loci such as the 18S rRNA gene (e.g., V1–V2, V4, or V9 variable regions, Baird & Hajibabaei, 2012; Bik, 2019; Geisen et al., 2018; Leasi et al., 2018; Lejzerowicz et al., 2015; Shaw et al., 2016; Thomsen & Willerslev, 2015), the 28S rRNA gene (e.g., D2–D3 domains, Pochon, Wecker, Stat, Berteaux-Lecellier, & Lecellier, 2019; Porazinska et al., 2009; Smith, Kohli, Murray, & Rhodes, 2017), and the Internal Transcribed Spacer regions (ITS1-ITS2, also encompassing the 5.8S rRNA gene, Lindner et al., 2013; Toju & Baba, 2018). Ribosomal RNA genes are easily amplified from a broad range of taxa using a variety of "universal" PCR primers, and the historical use of such genes in phylogenetic studies has resulted in substantial public sequence repositories including curated reference databases such as SILVA and PR2 (Guillou et al., 2013; Quast et al., 2013). Although rRNA loci typically undergo concerted evolution within eukaryotic genomes (whereby gene copies are homogenized across repeated rRNA arrays, Buckler, Ippolito, & Holtsford, 1997; Eickbush & Eickbush, 2007), divergent rRNA gene copies can persist within individuals, populations, and species (Bik, Fournier, Sung, Bergeron, & Thomas, 2013; Elbrecht, Vamos, Steinke, & Leese, 2018; Keller, Veltsos, & Nichols, 2008; Lindner & Banik, 2011; Pereira & Baldwin, 2016; Qiao et al., 2019). Thus, intragenomic rRNA variation can confound alpha- and beta-diversity estimations in metabarcoding studies, since polymorphic sequences usually result in multiple molecular operational taxonomic units (MOTUs) associated with each individual organism.

One major problem of eukaryotic metabarcoding studies is how to identify and circumvent the issue of intragenomic rRNA variation. Bioinformatic tools for copy correction of rRNA metabarcoding datasets have been developed exclusively for prokaryotes, since bacterial and archaeal species can also have multiple rRNA gene copies and exhibit some level of intragenomic polymorphism (Callahan et al., 2019; Sun, Jiang, Wu, & Zhou, 2013). Tools such as PICRUSt (Langille et al., 2013), CopyRighter (Angly et al., 2014), and PAPRICA (Bowman & Ducklow, 2015) rely on well-curated databases and phylogenetic frameworks to predict and correct for 16S rRNA copy number in bacterial and archaeal species. While many eukaryotic studies utilize alternate metabarcoding loci (e.g., mitochondrial genes with lower levels of intragenomic variation and higher taxonomic resolution at the species level; Kumar et al., 2017), the application of other gene regions is problematic in many metazoan groups such as marine nematodes which lack "universal" mitochondrial primer binding sites (Blaxter et al., 2005), and where the diversity of undescribed species is especially high (Blaxter, 2016). Thus, there is a pressing need to characterize patterns of rRNA variation across a broad range of microbial eukaryote taxa. Research efforts in this area are needed in order to elucidate the relationship between genomic patterns and traditional specimen-level data, identify correlations between molecular variation and ecology and life-history traits, and rapidly expand public database resources (thereby increasing the accuracy and precision of taxonomy assignments for eukaryotic MOTUs).

In the present study, we assessed patterns of 18S rRNA variation from individual marine nematode specimens, including potentially confounding effects on diversity estimations (e.g., relic DNA: DNA remaining in the marine sediment; gut contents: potential nematode prey items). Using a previously published 18S rRNA metabarcoding dataset generated from individual marine nematodes (Schuelke, Pereira, Hardy, & Bik, 2018), we estimated amplicon sequence variants (ASVs) using DADA2 (a software tool that models and corrects Illumina-sequenced amplicon errors, Callahan et al., 2016). Metabarcoding data were analyzed in conjunction with full-length reference sequences (~1,600 bp Sanger sequences of the 18S rRNA gene) generated for a subset of nematode morphospecies. Our investigation focused on three main questions: (a) From an evolutionary perspective, do nematodes with similar life histories share similar genomic patterns? We hypothesized that phylogenetically related nematodes (i.e., within the same genus or clade) would show similar ASV profiles or "Head-Tail" patterns of dominant and minor

MOTUs (Porazinska, Giblin-Davis, Esquivel, et al., 2010; Porazinska, Giblin-Davis, Sung, Thomas, & Others, 2010). (b) From an ecological perspective, do patterns of rRNA variation in marine nematode species correspond to known life-history traits? We hypothesized that predatory nematodes (feeding group 2B) would have more complex and diverse ASV profiles than bacterivorous nematodes (feeding group 1A). (c) Do we find consistency across different data types and methods of diversity estimation (i.e., classical morphological taxonomy, Sanger-based DNA barcoding, and high-throughput metabarcoding)? We hypothesized that rRNA polymorphisms observed in Sanger DNA barcodes would be also detected among ASVs and that morphological species would be largely congruent with molecular phylogenetic clade structures.

2 | MATERIALS AND METHODS

2.1 | 18S rRNA metabarcoding of individual marine nematode specimens

The 18S rRNA metabarcoding dataset derived from marine nematodes was previously generated as part of Schuelke et al. (2018). Individual marine nematode specimens were picked from sediment samples representing a diverse range of habitats and geographic regions (Table S1). These same nematodes, representing 44 marine nematode genera, were morphologically identified and imaged on temporary slide mounts under light microscopy (Nikon Eclipse E600; Nikon Corporation) prior to DNA extraction and PCR (i.e., one nematode per tube; Figure 1). Where possible, nematode specimens were separated into putative morphospecies (i.e., sp1, sp2, and so on; Table S1) when morphological variation within a genus was also supported by DNA sequence differences and clade structure observed in molecular phylogenies. Additional details on sediment sample collection and sample processing are provided in Schuelke et al. (2018). The nematode taxonomic classification adopted here follows the World Registers of Marine Species (WoRMS) database (http://www.marinespecies.org/index.php), whereas the feeding type classification is based on buccal (mouth) morphology groupings proposed by Wieser (1953): 1A—selective deposit feeders, 1B—nonselective deposit feeders, 2A—epigrowth feeders, and 2B—predators.

DNA extractions, PCR, and Illumina sequencing were carried out according to the methods in Schuelke et al. (2018). Briefly, individual nematodes were transferred into separate 0.2-ml PCR tubes following taxonomic identification, and DNA extractions were carried out using a Proteinase K "Worm Lysis Buffer" protocol where samples were incubated in a Thermomixer heated shaker block (Eppendorf) at 65°C and 750 rpm for 2 hr, followed by a 5-min incubation at 100°C to inactivate proteinase K. Although we did not include PCR replicates, each nematode morphospecies in our study was usually represented by multiple samples, so that we could also have some sort of replication at the species/clade level. Taxonomy blank samples (i.e., those containing only WLB, where the taxonomist simulated the transfer of nematodes into tubes) were also included as a checkpoint for potential sources of contamination (i.e., airborne microorganisms, reagent contaminants) in the laboratory. Lysates were used immediately or stored at −20°C. DNA extracts from individual nematodes were subsequently used to generate 18S rRNA metabarcoding datasets using the V1–V2 hypervariable region optimized for environmental amplification of microbial metazoan rRNA genes (F04 and R22 primers, Creer et al., 2010). Dual-index primer constructs were designed by modifying the Earth Microbiome Project Illumina amplicon protocol (http://www.earthmicrobiome.org/, Caporaso et al., 2012) and included a second barcode in the reverse primer. All primer constructs and oligo sequences have been made available on Figshare (https://doi.org/10.6084/m9.figshare.5701090). To avoid contamination, PCRs were set up in a dedicated laminar-flow hood that underwent daily sterilization with bleach and UV light. Individual nematode and blank/control sample PCRs had a final volume of 25 μl and contained 1 μl of DNA template, 0.5 μl of each primer (10 μM), 10 μl of Platinum Hot Start PCR Master Mix (2×) (Thermo Fisher), and 13 μl of molecular-grade water. Both negative (molecular-grade water) and positive (ZymoBIOMICS™ Microbial Community Standard; Zymo Research) controls were included in all PCRs.

All PCR products were purified using magnetic beads following the manufacturer's protocol (Agencourt AMPure XP beads; Beckman Coulter). Sample concentrations were subsequently measured using a Qubit® 3.0 Fluorometer and a Qubit® dsDNA HS (High Sensitivity) Assay Kit (Thermo Fisher Scientific). Normalization values were calculated to ensure that approximately equivalent DNA concentrations were pooled across all samples (including controls and blank samples). The final library was subjected to an additional magnetic bead cleanup step, followed by size selection on a BluePippin (Sage Science) to remove any remaining primer dimer and isolate target PCR amplicons within the range of 300–700 bp. A Bioanalyzer trace was run on the size-selected pool as a quality control measure, and the pooled 18S rRNA amplicon library was sequenced on an Illumina MiSeq platform (2 × 300-bp paired-end run) at the UC Davis Genomics Core Facility (Schuelke et al., 2018). All wet laboratory protocols and downstream bioinformatics scripts used in this study have been deposited on GitHub (https://github.com/BikLab/nematode-rRNA-variants).

2.2 | Generating nematode reference sequences via Sanger sequencing

We used Sanger sequencing to generate 18S rRNA barcodes from a subset of 86 nematode specimens included in the original study (Schuelke et al., 2018) where sufficient volumes of DNA template remained. We additionally isolated another 24 nematode specimens collected from the same Arctic samples (Table S1) in order to improve our 18S rRNA reference database and increase the accuracy of nematode taxonomic assignments derived from ASV sequences. The nearly full-length (~1,600 bp) 18S rRNA genes of 110 nematodes were amplified via PCR using three overlapping primer sets

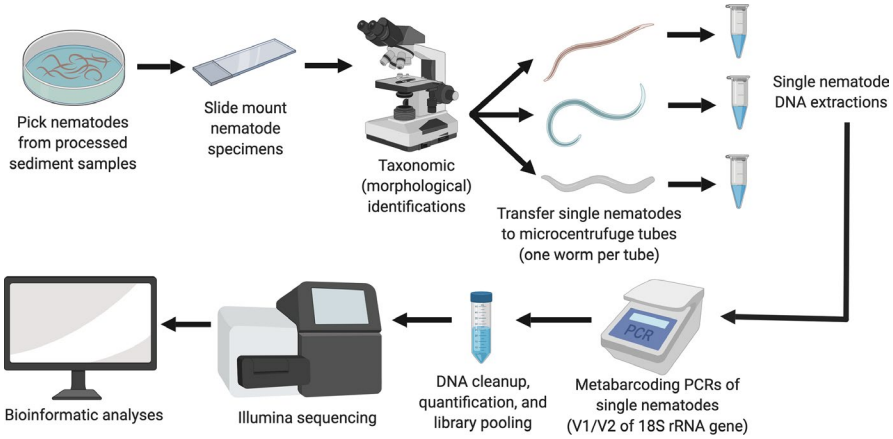

FIGURE 1 Workflow diagram of single-nematode metabarcoding. Free-living marine nematodes were isolated from processed marine sediment samples, mounted on glass slides, and identified down to genus level and/or putative morphospecies. DNA extractions and metabarcoding PCRs were subsequently carried out for single worms, followed by Illumina MiSeq sequencing and downstream bioinformatic analysis. Figure created with BioRender.com

(G18S4 and R26, 22F and 13R, and 24F1 and 18P, Bik, Lambshead, Thomas, & Lunt, 2010; Blaxter et al., 1998). All PCRs were 25 μl total volume, containing 3 μl of nematode DNA template, 1 μl of each primer (10 μM), 12 μl of Q5® Hot Start High-Fidelity 2× Master Mix (New England Biolabs), and 8 μl of molecular-grade water (HyClone HyPure Water, GE Healthcare Life Sciences). The following PCR profile was used for amplification of all 18S rRNA gene fragments: 98°C for 30 s; 98°C for 10 s, 55.4°C for 30 s, and 72°C for 30 s for 35 cycles; and 72°C for 2 min. Amplification success was evaluated via electrophoresis on a 1% agarose gel stained with SYBR® Green. Successful PCRs were purified with Agencourt AMPure XP beads (Beckman Coulter) using an in-house magnetic bead cleanup protocol. Finally, the 18S rRNA gene was sequenced in both directions with PCR primers using ABI-PRISM® Dye-DeoxyTerminator Big DyeTM v3.1 (Applied Biosystems) on an automatic sequencer Gene Analyzer® ABI 3100 (Applied Biosystems) at the Institute for Integrative Genome Biology, UC Riverside. Newly obtained 18S rRNA sequences were manually checked, edited, and assembled using CodonCode Aligner v. 4.2.7 (CodonCode Corporation, LI-COR, Inc.). DNA sequences representing the same nematode morphospecies were aligned using the MUSCLE algorithm implemented within CodonCode Aligner (Edgar, 2004) for evaluating the existence of intraspecific polymorphism. After removing primer sequences and carefully checking for ambiguous sites, high-quality contigs were exported for each nematode sample.

2.3 | Illumina data processing and generation of ASVs

Raw Illumina data were demultiplexed using a custom script for handling dual-index barcode combinations (available on GitHub). Next, the demultiplexed 18S rRNA dataset was analyzed in QIIME2 version 2019.10 (Bolyen et al., 2019) where primer sequences were trimmed using the cutadapt plugin (Martin, 2011). Denoising was based on optimal parameters (forward and reverse reads truncated at 232 and 253 bp, respectively, and a median PHRED score of ≥30). ASVs were subsequently generated using the DADA2 algorithm (Callahan et al., 2016). The estimation of ASVs is based on 100% sequence identity (or 1 nucleotide difference) and accounts for every single mutation in the dataset, thus allowing us to determine the frequency and importance of base pair changes in nematode rRNA gene copies. DADA2 was run using default parameters, including default chimera checking parameters (*consensus* option, which carries out de novo chimera identification and removes ASVs identified as chimeras if they are present in a significant fraction of samples). Taxonomy assignments for ASVs were obtained via the BLAST+ consensus taxonomy classifier (Camacho et al., 2009), using a custom reference database trimmed at the Illumina barcode length (~350 bp) and a minimum confidence value of 70%. Our custom reference database consisted of the QIIME-formatted SILVA 132 release (https://www.arb-silva.de/no_cache/download/archive/qiime/), a collection of marine nematode 18S rRNA barcodes produced by Macheriotou et al. (2019), and the nearly full-length 18S rRNA gene sequences generated from marine nematodes as part of this study (database files available on Github: https://github.com/BikLab/nematode-rRNA-variants).

2.4 | Bioinformatic analyses of ASV patterns and phylogeny reconstruction

Our final dataset differs slightly from that presented by Schuelke et al. (2018). We reduced the 18S rRNA metabarcoding dataset down to 227 individual nematode samples by removing samples with ambiguous nematode taxonomy (i.e., specimens identified to order level and above), relatively low-read counts (i.e., samples with <500 reads after DADA2 analysis), and very high-read counts (e.g., 4 samples had >100,000 reads). Thus, we have removed all outlier samples that could impact the interpretation of overall patterns.

The resulting ASV table was summarized and analyzed in order to assess patterns and sources of rRNA variation associated with individual marine nematodes. First, seven initial metrics were calculated for each nematode specimen (Table S2): the total number of demultiplexed and quality trimmed reads, the number of reads retained by DADA2, total number of ASVs, the number of ASVs with taxonomy assignments to Nematoda, the number of ASVs with ≥1% relative abundance that had taxonomy assignments to Nematoda, the relative abundance of the dominant ASV, and

TABLE 1 Number of nematode samples (also presented as a percentage of the total) belonging to the different sample categories. Classification is based on the most common ASV profiles (i.e., RA% of ASVs in a sample, see Section 2 for further details)

Sample categories	ASVs RA% range	Number of nematodes	Nematodes (%)
(1) Highly dominant ASV	ASV01 ≥85%	25	11.0
(2) Dominant ASV	ASV01 75%–84%	76	33.5
(3) No dominant ASV	ASV01 <75%	46	20.3
(a) Chloroplastida		6	13.0
(b) Fungi		13	28.3
(c) Nematoda		19	41.3
(d) Polychaeta		4	8.7
(e) Other[a]		4	8.7
(4) Intragenomic variation[b]	ASV01 <75% and	80	35.2
(a) 10%–20%	ASV2 ≥10%	67	83.8
(b) 21%–30%		10	12.5
(c) >30%		3	3.8

[a]Subcategory "Other" includes nematode samples having coamplification of SAR (2 nematode samples), Hydrozoa (1 sample), and Arachinada (1 sample).

[b]Nematode samples in the "intragenomic variation" category are further divided into ranges based on the RA% of the second most abundant ASV.

whether or not the dominant ASV corresponded to the expected nematode morphospecies. Nematode samples were also classified into four categories (Table 1) based on the observed ASV profiles produced by DADA2: (1) *Highly dominant ASV*, where the relative abundance of the dominant ASV was ≥85% and taxonomy matched the expected nematode ID; (2) *Dominant ASV*, where the relative abundance of the dominant ASV ranged from 75% to 84% and the taxonomy assignment matched the expected nematode ID; (3) *No dominant ASV*, where the relative abundance of the dominant ASV was <75% and the taxonomy assignment did not necessarily match the expected nematode ID; and (4) *Intragenomic variation*, where both the dominant and second most abundant ASV matched the nematode ID, and exhibited relative abundances of <75% and ≥10%, respectively. For some samples, the lack of a dominant or highly dominant ASV matching the expected nematode ID appeared to be related to the coamplification of DNA from other taxa (e.g., fungi, polychaetes, other nematode genera; Table S2 and discussion below).

Next, we explored the relationship between metabarcoding taxonomy assignments and ASV profiles obtained from individual nematodes. In particular, we evaluated patterns across samples representing the same nematode family, genus, phylogenetic clade (i.e., nematode specimens having the same dominant ASV or Sanger DNA barcode), and feeding group. The relationship between the number of ASVs and number of reads retained by DADA2 was estimated using Pearson's correlation coefficient with the package ggpubr v.0.2 in *R* v.3.6.0 (R Core Team, 2019). Data normality was assessed using Shapiro–Wilk's method, and Kruskal–Wallis (K-W) tests were used to assess differences among nematode trophic groups and nematode families with the package FSA v0.8.24 in *R* version 3.6.0 (R Core Team, 2019). The Mann–Whitney *U* test with adjustments for p-value (BH method; Benjamini & Hochberg, 1995) was used for pairwise comparisons (Zar, 2010). We also extracted diversity estimates, including Shannon diversity H' (Log_2), Margalef's species richness (*d*), and Inverted Simpson (D) diversity from the ASV table using PRIMER v7 software (Clarke & Gorley, 2015) and compared among nematode families and feeding groups. Ranked ASV distribution curves (i.e., number of reads vs. ranked ASV) were generated with ggplot2 v.3.1.1 in *R* (Wickham, 2016) to explore variation on ASV abundance derived from individual nematode specimens (and also averaged across nematode specimens representing the same nematode family) in order to identify common trends (e.g., Head-Tail patterns, Porazinska, Giblin-Davis, Sung, et al., 2010).

Detailed phylogenetic analyses of ASV profiles were also carried out for a subset of well-sampled nematode families and genera: Chromadoridae (Chromadoridae spp., *Chromadorella* spp., *Dichromadora* sp., *Euchromadora* sp., *Neochromadora* sp.), Comesomatidae (*Cervonema* spp., Comesomatidae spp., *Sabatieria* spp., and *Setosabatieria* sp.), Desmoscolecidae (*Desmoscolex* spp.), and Oxystominidae (*Halalaimus* spp., *Litinium* sp., *Oxystomina* spp., Oxystominidae spp., and *Thalassoalaimus* spp.). Phylogenetic analyses were performed using a fast maximum likelihood (ML), RAxML-HPC v.8 (Randomized Axelerated Maximum Likelihood, Stamatakis, 2014), through the server CIPRES (http://www.phylo.org/) under the GTR+G model as estimated in jModelTest (Darriba, Taboada, Doallo, & Posada, 2012). Gamma parameters were estimated from log-likelihood units, and bootstrap support was automatically calculated for the best scoring ML tree (Stamatakis, 2014). For each nematode specimen, we subsequently plotted the ASV profiles (i.e., relative abundance of all ASVs) as well as the taxonomic assignments of ASVs according to the following categories: (a) ASVs corresponding to the expected nematode morphospecies, (b) ASVs corresponding to other nematodes, and (c) ASVs corresponding to other non-nematode eukaryotes. All tree annotations, including plotting associated

metadata, were conducted using the package ggtree v.1.16.0 in *R* (Yu, Lam, Zhu, & Guan, 2018).

Similarly, we explored the ASV profiles of nematodes representing different feeding groups (i.e., 1A, 1B, 2A, and 2B) and tested whether or not the proportion of ASVs assigned to the three previous categories (i.e., expected nematode morphospecies, other nematodes, and other non-nematode eukaryotes) differs among these groups. Kruskal–Wallis and normality tests were carried out as described above. Next, we plotted the ASV profiles (number of reads and relative abundance of ASVs) of selected nematode specimens representing each feeding group (3 specimens per group) in a phylogenetic context. For this analysis, phylogenetic trees were built in MEGA 7 using the neighbor-joining method. Genetic divergence using p-distance (pairwise deletion of gaps/missing data) was estimated among sequences (Kumar, Stecher, & Tamura, 2016). Tree annotations were carried out using ggtree as described above (Yu et al., 2018).

3 | RESULTS

3.1 | Morphological and trophic diversity

A total of 24 families and 49 genera were identified in our marine nematode dataset (Figure S1, Table S1). Altogether, nematode families Comesomatidae (25.3%), Desmoscolecidae (12.3%), Oxystominidae (12.3%), and Chromadoridae (11.5%) represented about 61% of all identified nematode specimens. At the genus level, *Desmoscolex* (*n* = 32), *Sabatieria* (*n* = 28), *Halalaimus* (*n* = 17), *Dichromadora* (*n* = 15), and *Cervonema* (*n* = 13) were the most representative genera. The highest number of morphospecies, whose morphological variation was also supported by molecular data (i.e., these morphospecies represented different clades), was found in the genus *Desmoscolex* and *Sabatieria* with 13 and seven morphospecies, respectively (Figure S1). With respect to the trophic diversity, our dataset was mostly represented by selective deposit feeders (group 1A, *n* = 89) followed by nonselective deposit feeders (group 1B, *n* = 87) and epigrowth feeders (group 2A, *n* = 58), whereas predators (group 2B, *n* = 27) were relatively less abundant (Figure S1, Table S1).

3.2 | Diversity estimates with DADA2

Although the number of raw Illumina reads varied substantially across individually metabarcoded nematodes (1,346 to ~5.4 million), we observed that most samples (~38.4%) were in the 21,000–30,000 range (avg. = 55,466, median = 24,038). Error correction and ASV generation using the DADA2 algorithm further reduced these values (by 50% on average) and shifted the sample distribution: after DADA2, 45.1% of our nematode samples ended up in the 11,000–20,000 range (avg. = 33,753, median = 13,978). After further removing nematode outlier samples (i.e., 6 samples with low-read count, <500 reads; 4 samples with high-read count, >100,000 reads), read counts across nematode samples were much more consistent (avg. = 13,883, median = 14,035). Moreover, the stringency used on our DADA2 workflow drastically reduced the number of raw reads in some of the blank/negative control samples (Table S2).

The total number of ASVs obtained per sample ranged from 1 to 91 (avg. = 11, median = 9) and showed a significant but weak positive correlation with the number of retained reads (Figure 2a). Moreover, the number of ASVs did not vary significantly among nematode feeding groups (Kruskal–Wallis, $X^2 = 5.83$, $p = .12$) but was significant among nematode families (Figure 2b,c). However, pairwise comparisons with adjusted *p*-values (BH method) showed that these differences are likely to be a result of unbalanced sampling effort across nematode families (i.e., all pairwise comparisons had an adjusted $p > .05$). Diversity measures (e.g., Shannon [H'], Richness [d], and Simpson [D]) estimated from the ASV table showed similar patterns (Figure 3). For all three indexes, no significant differences were detected among nematode feeding groups (Kruskal–Wallis, $X^2 = 3.7$, $p = .29$ [H']; $X^2 = 5.83$, $p = .12$ [d]; $X^2 = 4.3$, $p = .23$ [D]), whereas some nematode families differed significantly ($p < .05$, after pairwise comparisons with adjusted *p*-values) for Shannon and Simpson indexes (Figure 3).

The Head-Tail pattern of rRNA variants associated with individual nematodes (ranked ASV curves; Figure 4) shows that the relative abundance of ASVs decreases very quickly and produces a "short tail" of rare ASVs with low-read counts. This was a very consistent trend across all four main nematode families, although specimens from the Oxystominidae displayed a slightly longer tail (Figure 4a). Overall, the dominant ASV for each nematode was about an order of magnitude more abundant than the second most abundant ASV. Lower variations in relative abundance were observed among less abundant ASVs (Figure 4b). We also observed that the relative abundance of the highest ranked ASVs for each worm was in the 81%–90% range and that only one nematode sample had an ASV with a relative abundance of 100% (e.g., Nem.227: *Halichoanolaimus* sp., Selachinematidae; Table S2).

3.3 | ASV profile patterns among nematode samples

To assess patterns of intragenomic rRNA variation in individual nematode specimens, we quantified the total number of ASVs per specimen (and the relative abundance of each ASV) with QIIME-derived assignments to nematode taxa. For 204 out of 227 nematode samples (90%), the ASV with the highest read count in each specimen's metabarcoding profile matched our expected morphological nematode genus ID (obtained via light microscopy). For the remaining 23 nematode samples (10%), the top ASV did not correspond to a nematode sequence (e.g., having a taxonomy assignment to a fungal or polychaete species), or alternatively, the top ASV was assigned to a different nematode genus than expected. However in both of the latter scenarios, our single-worm metabarcoding profiles always contained an ASV with the expected taxonomy assignment (the nematode genus ID determined via light microscopy), and

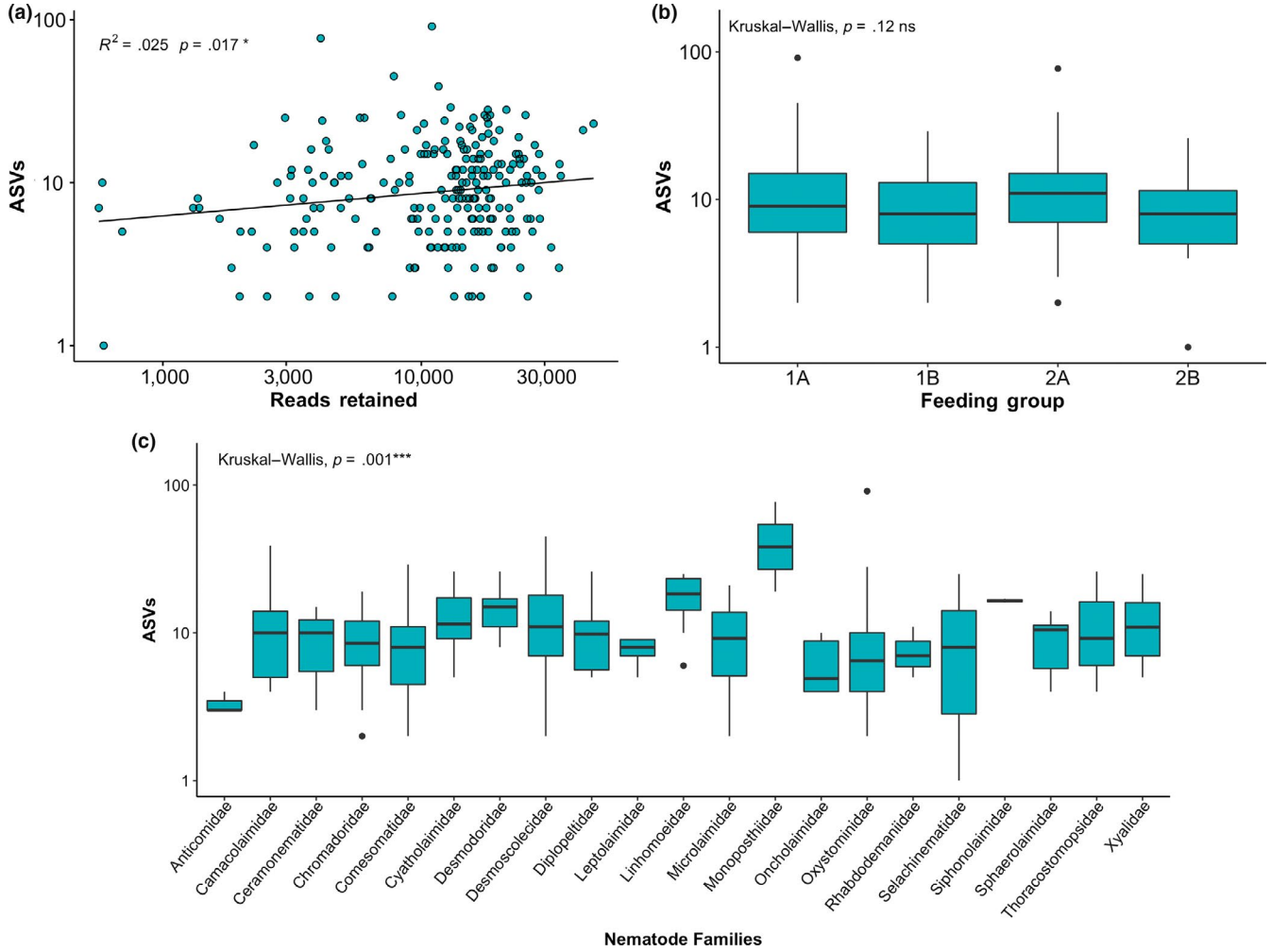

FIGURE 2 Delimitation of ASVs with the DADA2 method. (a) Relationship between number of observations per single-nematode sample (total number of ASVs) and number of reads retained by DADA2. (b) Number of ASVs recovered across different nematode feeding groups (1A—selective deposit feeders, 1B—nonselective deposit feeders, 2A—epigrowth feeders, and 2B—predators sensu Wieser, 1953). (c) Number of ASVs recovered across different nematode families

these species-barcode ASVs were always present at >1% relative abundance.

Based on the relative abundance profiles of all per-specimen ASVs with taxonomic assignments to nematodes, our samples were most commonly shown to have a "highly dominant" or "dominant" ASV present at >75% relative abundance in the overall profile (44% of specimens; Table 1). The second most common category was that of "intragenomic variation" where two competing nematode ASVs had <75% relative abundance in the specimen metabarcoding profile (35% of specimens; Table 1); often these two ASVs had relative abundance similar to each other and only differed by 10%–20% in terms of their overall relative abundance in the specimen metabarcoding profile. Moreover, half of the samples in the intragenomic variation category had at least three ASVs with a relative abundance ≥10%. The third most common category (20% of specimens, Table 1) was "no dominant ASV" mostly due to the coamplification of nontarget DNA, including other nematodes and fungi. Still within this category, further analysis of those samples where the most abundant ASV did not match the expected nematode (i.e., 23 out 46 samples) revealed that these ASVs were usually unique to a sample and not shared across the dataset. These findings suggest that the existence of other organisms' DNA in the samples most likely represents a real biological phenomenon, as opposed to sample contamination and/or a sequencing artifact derived from the Illumina platform (e.g., tag jumping; see discussion for further detail).

3.4 | ASV profile patterns among taxonomic/phylogenetic and feeding groups

Phylogenetic analysis allowed us to assess whether or not closely related nematode species share similar ASV profiles (e.g., same dominant ASV and ASVs with similar relative abundance values). Overall, ASV profiles were highly consistent across nematode specimens representing the same nematode family (Chromadoridae, Comesomatidae, Desmoscolecidae, and Oxystominidae; Figure 5 and Figure S2). Within the four nematode families we assessed in-depth, ASV patterns seemed to be clade specific—in other words,

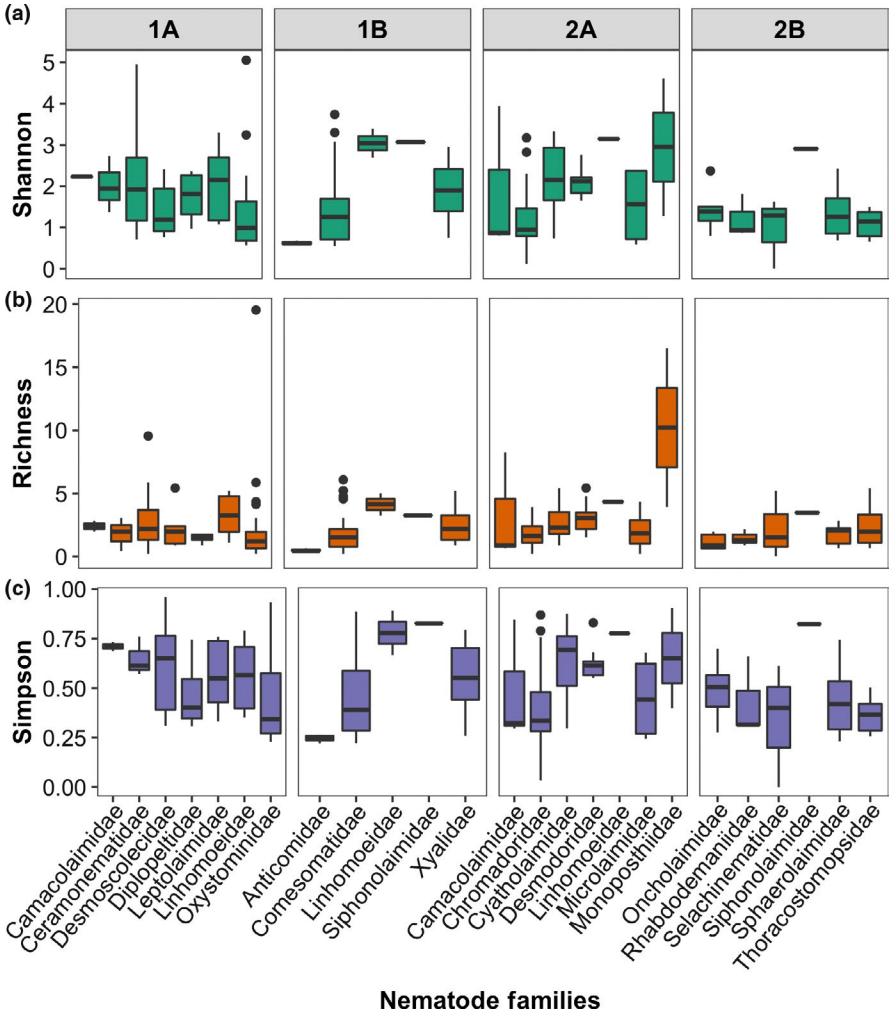

FIGURE 3 Diversity metrics according to nematode family and nematode feeding group. (a) Shannon diversity H' (Log_2). (b) Margalef's species richness (d). (c) Inverted Simpson (D) diversity index. Among nematode families, significant differences ($p < .05$) for Shannon included the following: Desmoscolecidae versus Comesomatidae, Thoracostomopsidade; Comesomatidae versus Desmodoridae; and Diplopeltidae versus Oxystominidae. In addition to those, for the Simpson index significant differences included the following: Desmoscolecidae versus Camacolaimidae, Cyatholaimidae

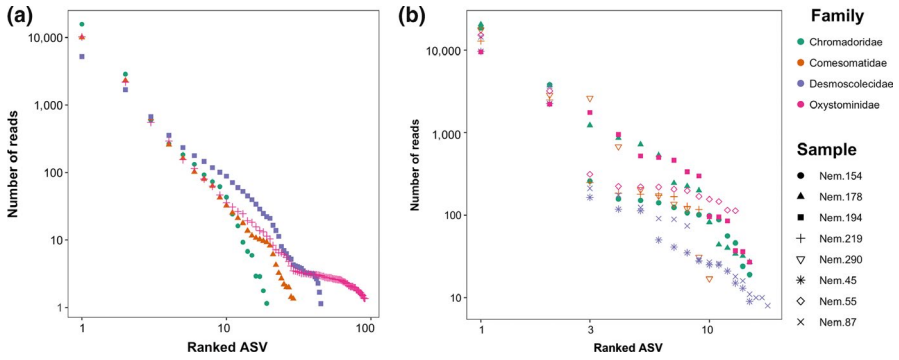

FIGURE 4 Ranked ASV distribution showing Head-Tail patterns. (a) Average ranked ASV distribution for the four main nematode families clustered in DADA2 display a truncated Head-Tail pattern with fewer overall ASVs. In this case, the number of reads (y-axis) are presented as adjusted mean values (i.e., mean + 1) for each ASV due to the logarithmic (Log_{10}) scale. (b) The same ranked ASV pattern is recovered from individual nematode samples (two specimens per family). Taxonomy of nematode samples is as follows: Chromadoriade, Nem.154 *Dichromadora* sp. and Nem.178 Chromadoridae sp4; Comesomatidae, Nem.219 *Cervonema* sp2, and Nem.290 Comesomatidae sp4; Desmoscolecidae, Nem.45 *Desmoscolex* sp1, and Nem.87 *Desmoscolex* sp5; and Oxystominidae, Nem.55 *Halalaimus* sp8, and Nem.194 Oxystominidae sp2

nematode specimens representing the same morphospecies and/or phylogenetic clade tended to have very similar ASV profiles. For example, specimens representing *Dichromadora* sp. (Chromadoridae), *Sabatieria* sp5 (Comesomatidae), and *Halalaimus* sp4 (Oxystomonidae) displayed almost identical ASV profiles (i.e., same top ASV sequences with similar relative abundance) within their respective clades (Figure 5b). Although ASV dominance patterns and taxonomic assignments of top ASVs seem to be mostly consistent

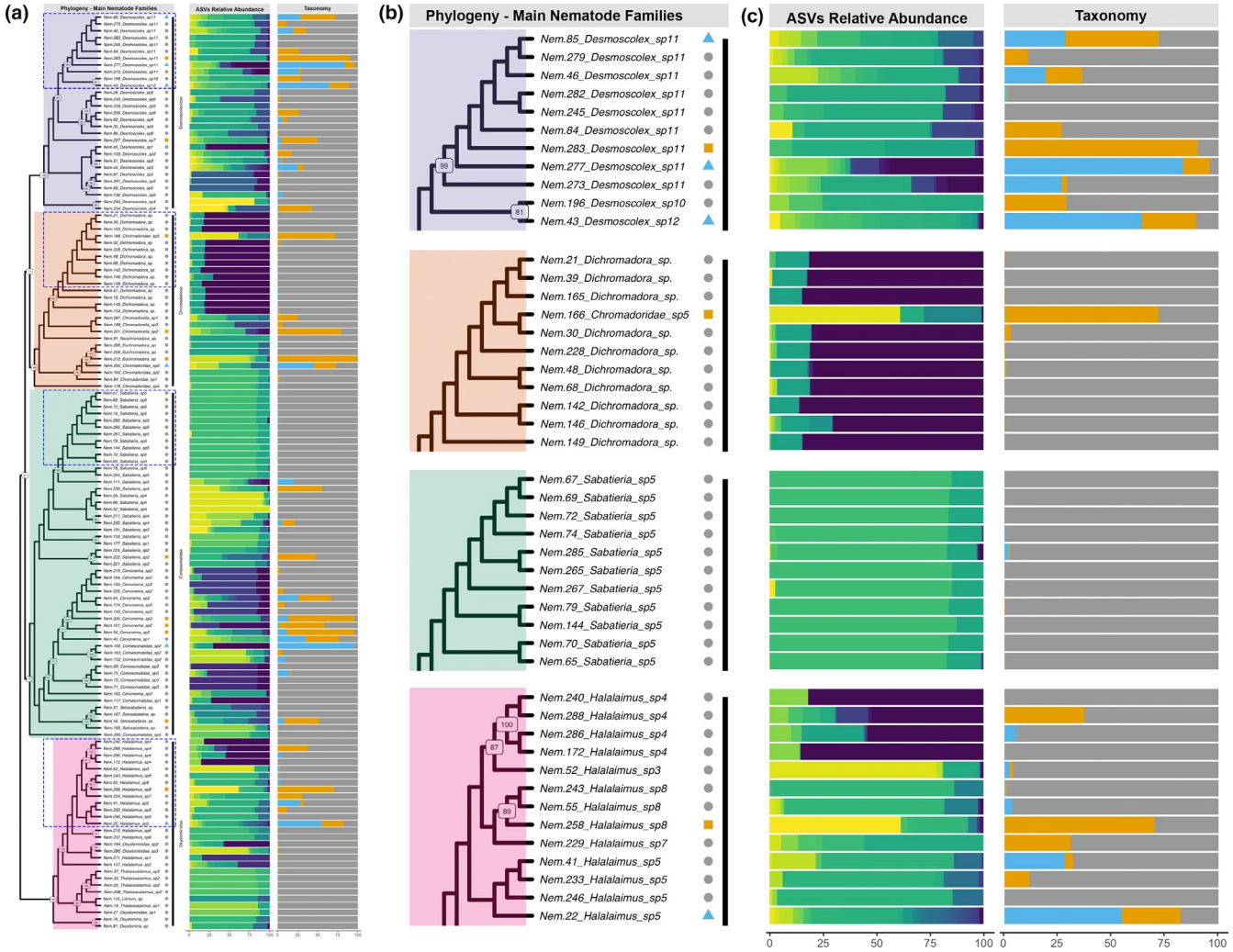

FIGURE 5 Composition and variation of DADA2 ASV profiles recovered across different phylogenetic clades representing four nematode families: Chromadoridae (orange), Comesometidae (green), Desmoscolecidae (purple), and Oxystominidae (pink). On each branch, the taxonomic assignment of the dominant ASV recovered from each nematode specimen is indicated as the correct nematode morphospecies (gray circle), another nematode (blue triangle), or a non-nematode eukaryote (orange square). (a) Maximum-likelihood tree reconstruction based on the representative ASV sequence (i.e., matching the morphological nematode ID). Tree is rooted on the branch leading to all representatives of the family Oxystominidae (BS > = 70% are shown). (b) ASV profiles based on the relative abundance of all ASVs (presented in a continuous color-coded scale) for selected nematodes specimens, indicated by dotted blue boxes in panel a. (c) Taxonomic assignments of ASVs are summarized by color: ASVs corresponding to the expected nematode morphospecies (gray), other nematodes (blue), and other non-nematode eukaryotes (orange). A complete ML tree including all nematode specimens representing these four main families is provided in Figure S2

within taxonomic/phylogenetic groups, there were also some cases where ASV profiles varied (e.g., Nem.283 *Desmoscolex* sp11, Nem. 166 Chromadoriade sp5). Variation in ASV profiles usually occurred when the dominant ASV did match a different nematode group or another eukaryotic taxon (e.g., fungi, polychaete) instead of the expected nematode genus ID. When comparing the overall ASV taxonomy associated with these inconsistent nematode samples, we also observed a higher proportion of ASVs matching groups other than the nematode morphospecies (Figure S2), suggesting higher coamplification of nontarget eukaryotic taxa. Moreover, nematode specimens with higher putative levels of coamplification tended to have a higher number of ASVs compared to other members of the same clade.

With respect to trophic diversity, we did not find significant differences in ASV profiles among nematode feeding groups (i.e., 1A, 1B, 2A, and 2B); the total number of reads (data not shown) and the number of ASVs did not vary significantly across these four groups (Figure 2b). Similarly, when ASVs are categorized based on their taxonomic assignment (i.e., expected nematode morphospecies, other nematodes, and other non-nematode eukaryotes), the pattern is very consistent across nematode feeding groups. As expected, there was a significantly higher proportion of ASVs matching the expected nematode morphospecies (Figure S3), whereas no significant differences are observed between ASVs assigned to other nematodes and non-nematode eukaryotes (except for feeding group 1A, where poorly sampled nematode clades may

explain this statistical result). On the other hand, variation among ASV categories may show a correlation with nematode taxonomy, including variations within clades (e.g., among closely related species and even among nematode specimens representing the same morphospecies; Figure S3).

3.5 | Intragenomic variation as a source of rRNA patterns

Nematode specimens exhibiting intragenomic variation represented over one-third of the samples in our dataset (Table 1). In this sample category, nematode specimens were represented by at least two abundant ASVs with taxonomy of both ASVs matching the expected nematode morphospecies. Overall, the relative abundance decreased considerably from the first to the second most abundant ASV, although there were also cases where both ASVs showed comparable abundances (e.g., Nem.51 *Setosabatieria* sp.: ASV1 = 52%, ASV2 = 38%; Nem.38 *Halichoanolaimus* sp.: ASV1 = 52%, ASV2 = 34%). Moreover, all nematode samples belonging to the category of "dominant ASV" also had another ASV matching the expected nematode species (and present at ≥11% relative abundance), suggesting that Head-Tail patterns of ASVs and multiple intragenomic rRNA variants exist in the majority of free-living marine nematode species (~70% of the samples we examined in this study). When variation among ASVs was compared at the sequence level, we also observed that intragenomic variation, as estimated by p-distance, can vary drastically among nematode morphospecies (Figure S4, Table S3). In some nematodes, intragenomic rRNA variants diverged by as much as 5%–20% sequence identity. However, the location of polymorphic sites (observed in multiple sequence alignments) appeared to be clade specific. As an example, nematode specimens representing *Dichromodora* sp. had highly divergent ASVs (p-distance: mean = 2.57%, range = 0%–20.95%). However, we observed that the Sanger-derived nematode barcode (i.e., the reference sequence for a morphospecies) exhibited 100% pairwise identity to the dominant ASV in most of the cases (Figure S4), suggesting that the dominant ASV most likely represents the rRNA variant sequence with the highest copy number within a nematode genome.

4 | DISCUSSION

4.1 | Consistent levels of rRNA variation across individual nematode specimens

Metabarcoding studies focusing on prokaryotic and eukaryotic diversity have heavily relied on rRNA genes owing to both theoretical and practical reasons (Bik, Porazinska, et al., 2012; Bik, Sung, et al., 2012; Deiner et al., 2017; Ibal, Pham, Park, & Shin, 2019; Leasi et al., 2018; Lindner et al., 2013). However, the multicopy nature of rRNA genes and the existence of intragenomic variation among these gene copies confound our ability to accurately survey biodiversity and understand patterns and processes in nature. Numerous studies, whether using cloning techniques coupled with Sanger sequencing or newer HTS technologies such as Illumina, have extensively documented intragenomic rRNA variation in prokaryotes and eukaryotes (Alanagreh, Pegg, Harikumar, & Buchheim, 2017; Brabec, Kuchta, Scholz, & Littlewood, 2016; Lindner & Banik, 2011; Nieto Feliner, Gutiérrez Larena, & Fuertes Aguilar, 2004; Pereira & Baldwin, 2016; Sun et al., 2013). A common feature observed in eukaryotic rRNA datasets is the existence of a "Head-Tail" pattern as defined by Porazinska, Giblin-Davis, Esquivel, et al. (2010), whereby the "Head" represents a highly dominant MOTU (i.e., containing the majority of sequence reads) and the "Tail" comprises a set of rarer MOTUs each containing a decreasing number of reads.

In the present study, we explored 18S rRNA variation in marine nematodes from different perspectives, including assessment of both intragenomic variation (i.e., within-individual) and intraspecific variation (i.e., among specimens representing the same nematode morphospecies). Our results showed that Head-Tail patterns were a common and consistent phenomenon among individual nematode specimens and all major nematode families. Moreover, Head-Tail patterns in our dataset were persistent even when using the newest bioinformatic algorithms that employ error-correction strategies to generate amplicon sequence variants (ASVs), which essentially create MOTUs clustered at 100% sequence identity. We observed that ASV pipelines such as DADA2 return a smaller number of ASVs associated with individual nematode specimens, as well as a much shorter tail of rare MOTUs compared to other methods (e.g., VSearch, data not shown).

Bik et al. (2013) estimated intragenomic rRNA copy number for six nematode species and reported values ranging from 56 to 323 rRNA copies, while the model nematode *Caenorhabditis elegans* is estimated to harbor 100–150 rRNA copies (Stein et al., 2003). In this study, we consistently recovered <100 nematode ASVs associated with individual worms (Figure 2), confirming that the DADA2 pipeline produces outputs that are consistent with known biological patterns of intragenomic rRNA variation in nematodes (Porazinska, Giblin-Davis, Esquivel, et al., 2010; Porazinska et al., 2009). Head-Tail patterns in our nematode samples did not appear to be influenced by sequencing depth, exhibiting only a weak correlation between ASVs and the number of reads (Figure 2a,b). In addition to providing high resolution between rRNA variants (i.e., differences of one nucleotide), DADA2 has been shown to outperform other methods by recovering more real variants and fewer spurious sequences (Callahan et al., 2016). Previous studies of marine nematodes also indicate that DADA2 ASVs produce the most realistic estimates of overall species richness, accurately reporting 30 out of 39 nematode species in a mock community assessment (Macheriotou et al., 2019). Thus, DADA2 ASVs are able to consistently recover variant rRNA copies that persist within eukaryotic genomes while simultaneously providing an accurate view of community assemblage structure. However, quantifying the exact number of intragenomic rRNA variants using metabarcoding datasets remains challenging, even when using ASV profiles resolved at the level of a single specimen.

4.2 | Clade-specific patterns of rRNA variants and recovery of nontarget taxa

Unsurprisingly, most of the dominant ASVs recovered in our dataset were associated with the expected nematode morphospecies, genus, or family. Although the number of 18S rRNA reads, and consequently the number of ASVs, varied across nematode specimens (Table S2), we observed that ASV patterns across nematode taxa are most likely to be clade specific (Figure 5, Figure S3). Individual nematodes representing the same morphospecies overall shared the same ASV nucleotide sequences (both dominant and minor rRNA variants), and these ASVs were recovered at very similar relative abundances across replicate worms from the same morphospecies, thus resulting into almost identical ASV profiles (i.e., ASV richness and diversity; see *Dichromadora* sp. and *Sabatieria* sp5 in Figure 5). However, the taxonomic composition of ASV profiles could also be impacted by other sources of eDNA which are coextracted and coamplified from single worms during wet laboratory protocols. For instance, in cases where the dominant ("Head") ASV did not match the expected nematode morphospecies, the ASV profile of those specimens differed substantially from other worms in the same phylogenetic clade. Coamplification of non-nematode eukaryotic DNA (e.g., fungi, arthropods, alveolates) from nematode samples has been previously reported by many molecular studies utilizing Sanger sequencing and cloning techniques (Bhadury & Austen, 2010; Bhadury et al., 2006, 2011; Bhadury, Bridge, Austen, Bilton, & Smerdon, 2009).

Newer HTS technologies are highly sensitive in detecting very low amounts of DNA, and the recovery of nontarget organisms (especially when using "universal" metabarcoding PCR primers) could lead to misinterpretations of microbial eukaryote assemblage structure, especially when generating metabarcoding data from bulk sediments. An alternative explanation for those cases where the dominant ASV did not match the expected morphospecies (especially when the recovered match was a different nematode genus) is the issue of "tag jumping" in Illumina-based metabarcoding studies, where reads may be mis-assigned to samples during demultiplexing and obscure diversity patterns (Schnell, Bohmann, & Gilbert, 2015). Errors (i.e., false positives) due to tag jumping are generally reported to be relatively low: 0.01%–0.03% specifically for phiX (Hänfling et al., 2016; Olds et al., 2016); or 2.1%–2.6% within specific libraries (Schnell et al., 2015). However, these levels will certainly vary among metabarcoding studies depending on the sequencing strategy adopted (single- versus paired-end barcoding, Illumina platform used, etc.). Although we cannot entirely rule out the effects of tag jumping in this study, the consistency of our results suggested that, if present, its impact was minimum.

Recovery of some types of nontarget DNA can provide novel scientific insights, especially if this information is explicitly linked with specimen-level data. For example, many nematodes and microbial metazoan species are known to harbor microbial symbionts and parasites, such as microsporidian taxa infecting nematodes in both soil and marine environments (Gibson & Morran, 2017; Sapir et al., 2014). One recent metabarcoding study also reported distinct community profiles of protist taxa associated with metazoan DNA extracts (dominated by parasitic and pathogenic species, Geisen, Laros, Vizcaíno, Bonkowski, & de Groot, 2015) compared to protist mock community assemblages, indicating that host-associated microbial pathogens remain poorly characterized despite their seemingly common occurrence in natural habitats worldwide. However, it is important to note that coamplification of nontarget taxa does not necessarily imply ecological interactions, especially since most metabarcoding approaches cannot differentiate between living species (extant microbial communities) and relic DNA (dead organisms and extracellular DNA).

In the present dataset, some nematode groups such as the family Desmoscolecidae (Figure 5b) appeared to have a higher proportion of ASVs with taxonomy assignments to other nematodes and non-nematode eukaryotes. Morphologically, Desmoscolecidae are small nematodes (often <500 μm) characterized by short, oval-shaped bodies with conspicuous transverse rings also known as "desmen," which are composed of secretions and sediment particles including grains of sand (Decraemer & Rho, 2013; Platt & Warwick, 1988). Marine sediment particles are known to readily bind DNA and shield nucleotides from degradation (Torti, Lever, & Jørgensen, 2015), and the persistent attachment of sediment particles on the cuticle of Desmoscolecid nematodes may greatly facilitate the coamplification of relic DNA from the environment. Similarly, we noticed that some nematode specimens in the family Comesomatidae (e.g., *Cervonema* sp2) also displayed a large number of ASVs with taxonomic assignments to other nematodes and other non-nematode eukaryotes. Contrary to Desmoscolecidae nematodes, representatives of the Comesomatidae are much larger with a fusiform body shape and a smooth cuticle. Based on their buccal cavity morphology, this group is classified as nonselective deposit feeders (i.e., feeding group 1B sensu Wieser, 1953) and therefore is likely to ingest larger sediment particles and potentially relic DNA from the environment. Comesomatidae nematodes have been observed to prey on other small nematodes (Moens & Vincx, 1997) as well as being attracted to dead animals (Gerlach, 1977). Our single-nematode metabarcoding approach could not unequivocally determine whether recovery of nontarget taxa represented false positives due to tag jumping (see above), bioinformatic artifacts (e.g., inaccurate taxonomy assignments resulting from sparse rRNA databases), chance amplification of relic DNA, or active feeding or behavioral strategies of each nematode species. The coamplification of nontarget DNA is also impacted by the overall integrity of the specimen: Degraded nematode specimens and low-quality host DNA may increase the chance of amplifying microbial associates, relic DNA, and gut contents, as well as be impacted by potential primer biases which may promote recovery of certain taxa over others (Lanner, Curto, Pachinger, Neumüller, & Meimberg, 2019; Sow et al., 2019).

Our assessment of nontarget nematode taxa is also heavily impacted by database size, completeness, and the bioinformatic algorithm and parameters used to determine taxonomy assignments from metabarcoding ASVs (Macheriotou et al., 2019). In the present

dataset, the accuracy of taxonomic assignments for single-nematode ASV profiles appears to vary greatly depending on phylogenetic distance and the number of closely related 18S rRNA sequences present in our reference databases (SILVA and our local custom Sanger database of nematode specimens). For example, some of our poorly sampled morphospecies including *Halichoanolaimus* sp. (Nem.17, Nem.38, Nem. 227), *Neodiploteltula* sp. (Nem. 29), and *Pselionema* sp. (Nem.269, Nem. 274, Nem. 276) were either erroneously assigned to a nematode genus that was well-represented in our database (e.g., to *Sabatieria*, a genus where many full-length DNA barcodes are available) or only assigned to higher taxonomic ranks (i.e., order or class). However, we were able to circumvent some database issues by using phylogenetic analysis of 18S rRNA data to confirm placement of these morphospecies in the correct clades.

4.3 | No correlation between rRNA variation and nematode life-history traits

Although different nematode feeding strategies could potentially result in distinct ASV profiles, we did not depict a consistent pattern across our 18S rRNA dataset. The number of ASVs did not significantly differ among nematode feeding groups (1A, 1B, 2A, 2B sensu, Wieser, 1953; Figure 2c). Moreover, when we examined the taxonomic assignment of ASVs (i.e., matching the expected nematode morphospecies, another nematode, or a non-nematode eukaryote), we observed a high degree of variation within feeding groups (Figure S3). The majority of high-abundance ASVs matched the expected nematode morphospecies, and these sequences were always significantly different from ASVs matching other nematodes and non-nematode eukaryotes. These results further support the idea that ASV profiles are likely to correspond more strongly to nematode phylogenetic clades rather than (potentially arbitrary) ecological classifications such as feeding groups (Moens, Bouillon, & Gallucci, 2005).

We initially hypothesized that predatory nematodes (i.e., feeding group 2B) would have more complex ASV profiles compared to other nematode groups, owing to higher prey diversity, larger and more complex buccal cavities, and active feeding strategies (Jensen, 1987). This hypothesis was readily disproven by the present dataset, but we did recover some potential instances of potential predator–prey interactions. For one specimen in the family Sphaerolaimidae (Nem.231, *Subsphaerolaimus* sp1), we observed that the dominant ASV (69% abundance) did not match the expected nematode morphospecies, but instead a possible prey item (another nematode, Xyalidae sp.), whereas the second most abundant ASV (16% abundance) matched the expected nematode morphospecies. Additionally, some predatory nematodes representing Oncholaimidae (Nem. 101, *Viscosia* sp4) and Rhabdodemaniidae (Nem. 230, *Rhabdodemania* sp2) had the second most abundant ASV matching potential prey items (e.g., other nematodes, algae, with relative abundance always >10%). Overall, the complexity and structure of ASV profiles recovered from single-nematode specimens could not be attributed to any single factor. Patterns of intragenomic rRNA variation, recovery of nontarget taxa, and Head-Tail patterns of ASVs varied across different nematode genera. Further work is needed to determine whether some observed patterns may be specific to certain nematode species, populations, or geographic regions.

4.4 | Integration of data types helps clarify and circumvent intragenomic rRNA variation

In this study, we integrated traditional morphology-based taxonomy and molecular approaches, including Sanger sequencing and HTS, to evaluate the variation of 18S rRNA sequences associated with individual marine nematode specimens. High-resolution morphological taxonomy (i.e., at the genus or species level) of microbial metazoa such as nematodes is challenging, time consuming, and requires specialized taxonomic expertise (Bik, Porazinska, et al., 2012; Bik, Sung, et al., 2012; De Ley et al., 2005; Holovachov, Haenel, Bourlat, & Jondelius, 2017). Here, our integrated morphological–molecular approach allowed us to take into account all lines of evidence to delimit species and assess overall levels of biodiversity. "Reverse taxonomy" was often the most useful approach (Kanzaki et al., 2012; Markmann & Tautz, 2005), where the variation in metabarcoding ASVs and Sanger-derived 18S rRNA sequences guided our reassessment of morphological variation and helped to refine our initial nematode identifications.

For example, in the family Desmoscolecidae, we were able to identify 13 putative morphospecies, whose variation at the molecular level was also supported by morphological differences previously overlooked during our initial light microscopy observations. Similarly, a large number of putative nematode morphospecies were also identified in the three other major families viz. Comesomatidae, Chromadoridae, and Oxystominidae. Our results also showed that metabarcoding data strongly agree with Sanger sequencing data. For example, in most cases the Sanger-derived 18S rRNA barcode of nematode specimens was 100% identical to the dominant ASV sequence recovered from the metabarcoding dataset. This suggests that in studies of nematodes, the species-specific DNA barcode will be recovered as the dominant ASV and with high relative abundance. Moreover, using this short 18S rRNA fragment (i.e., V1–V2 region), we were able to detect mutations among closely related taxa that also supported morphological variation observed under light microscopy. Overall, our results showed high congruence across all of the applied methods.

The idea of supplementing HTS metabarcoding studies with additional datasets (e.g., morphological data) varies among eukaryotic groups, but it has been lately supported by many authors (Cahill et al., 2018; Dell'Anno, Carugati, Corinaldesi, Riccioni, & Danovaro, 2015; Geisen et al., 2018; Lejzerowicz et al., 2015; Macheriotou et al., 2019). These studies have highlighted the importance of classical taxonomy in characterizing biodiversity as well as advocating for the integration of both approaches, particularly for those groups poorly and/or sparsely represented in molecular databases

such as free-living marine nematodes. For example, V1–V2 regions of the 18S rRNA are commonly used for marine nematodes (Bik, Porazinska, et al., 2012; Bik, Sung, et al., 2012; Fonseca et al., 2010), whereas studies on terrestrial nematodes have targeted the V9 region (Porazinska, Giblin-Davis, Esquivel, et al., 2010; Porazinska et al., 2009). Moreover, many marine nematode groups lack formal taxonomic descriptions despite high levels of biodiversity (Miljutin et al., 2010), with many major groups having no representation at all in molecular databases. Thus, delimiting free-living nematode species and quantifying abundance from rRNA metabarcoding reads alone will continue to be fraught with difficulties unless more rapid progress is made in filling taxonomic and database gaps.

5 | CONCLUSIONS

The present study outlined patterns of intragenomic rRNA variation associated with individual free-living marine nematodes, providing assessments of ASV profiles across nematode phylogenetic clades and ecological feeding groups. While Head-Tail patterns and relative abundance of rRNA variants were consistent at the level of morphospecies (i.e., ASVs were recovered at very similar relative abundances across replicate worms, or higher levels of nontarget taxa were recovered some groups such as the sediment-encrusted Desmoscolecidae), the level and categorization of intragenomic rRNA variation (Table 1) appeared to vary widely across taxa. Interpretation of broad patterns was additionally confounded by the uncertainties related to ASV taxonomic assignments and sparse eukaryotic databases. Despite our use of labor-intensive wet laboratory protocols to isolate individual nematode specimens and minimize environmental DNA contamination, we consistently recovered signals from relic DNA and potential prey/parasite taxa associated with single nematodes. However, in most samples the signal from the target nematode morphospecies represented the dominant sequence ("Head") in the ASV profiles of individual worms, indicating that metabarcoding protocols will reliably report the species-specific DNA barcode, even in the presence of intragenomic rRNA variants and low levels of eDNA derived from other sources. Furthermore, amplification of nontarget taxa can provide novel ecological insights on host–microbe interactions and feeding strategies, generating observational data and hypotheses that can be explored in future studies of microbial metazoa. The choice of metabarcoding marker locus (e.g., rRNA loci vs. mitochondrial genes), reference database composition and completeness, and read clustering and taxonomy assignment algorithms will continue to have a large impact on the assessment of biodiversity patterns and species presence/absence in HTS datasets (Holovachov et al., 2017; Macheriotou et al., 2019). However, the use of complementary datasets (environmental metabarcoding data and specimen-level morphology and reference DNA barcodes) can provide a robust and accurate view of microbial eukaryote community assemblages, particularly for rRNA gene loci where pervasive intragenomic variation confounds our ability to quantitatively link -Omics data with individual specimens.

ACKNOWLEDGMENTS

This research was made possible by a grant from The Gulf of Mexico Research Initiative. Data are publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org (doi: 10.7266/N7959G4X). Additional funding for this was provided by a North Pacific Research Board award to SH and HMB (NPRB project 1303).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

TJP and TS picked and isolated nematode specimens, carried out taxonomic identifications and wet laboratory molecular protocols, and generated Illumina and Sanger sequencing data. TJP, ADS, and HMB carried out data analysis, interpreted results, and wrote the manuscript. SMH collected marine sediment samples, and HMB and SMH contributed to the conception and design of this study.

DATA AVAILABILITY STATEMENT

Raw Illumina metabarcoding data (18S rRNA gene amplicons generated from single nematodes) have been deposited in the NCBI Sequence Read Archive (BioProject PRJNA422296 and SRA accession SRP128131). Full-length 18S rRNA gene sequences newly generated via Sanger sequencing have been deposited on GenBank (Accession Numbers: MN250033–MN250142) Metabarcoding primer constructs and QIIME mapping files are available on Figshare (data from Schuelke et al., 2018; https://doi.org/10.6084/m9.figshare.5701090). QIIME outputs (ASV tables), and all scripts used for processing and analyzing the data in the present study are available on GitHub (https://github.com/BikLab/nematode-rRNA-variants).

ORCID

Tiago José Pereira https://orcid.org/0000-0002-6424-2848
Alejandro De Santiago https://orcid.org/0000-0001-9086-3050
Holly M. Bik https://orcid.org/0000-0002-4356-3837

REFERENCES

Alanagreh, L., Pegg, C., Harikumar, A., & Buchheim, M. (2017). Assessing intragenomic variation of the internal transcribed spacer two: Adapting the Illumina metagenomics protocol. *PLoS ONE*, *12*(7), e0181491. https://doi.org/10.1371/journal.pone.0181491

Angly, F. E., Dennis, P. G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P., & Tyson, G. W. (2014). CopyRighter: A rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome*, *2*, 11. https://doi.org/10.1186/2049-2618-2-11

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8), 2039–2044.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bhadury, P., & Austen, M. C. (2010). Barcoding marine nematodes: An improved set of nematode 18S rRNA primers to overcome

eukaryotic co-interference. *Hydrobiologia*, *641*(1), 245–251. https://doi.org/10.1007/s10750-009-0088-z

Bhadury, P., Austen, M. C., Bilton, D. T., Lambshead, P. J. D., Rogers, A. D., & Smerdon, G. R. (2006). Molecular detection of marine nematodes from environmental samples: Overcoming eukaryotic interference. *Aquatic Microbial Ecology*, *44*(1), 97–103. https://doi.org/10.3354/ame044097

Bhadury, P., Bik, H., Lambshead, J. D., Austen, M. C., Smerdon, G. R., & Rogers, A. D. (2011). Molecular diversity of fungal phylotypes co-amplified alongside nematodes from coastal and deep-sea marine environments. *PLoS ONE*, *6*(10), e26445. https://doi.org/10.1371/journal.pone.0026445

Bhadury, P., Bridge, P. D., Austen, M. C., Bilton, D. T., & Smerdon, G. R. (2009). Detection of fungal 18S rRNA sequences in conjunction with marine nematode 18S rRNA amplicons. *Aquatic Biology*, *5*(2), 149–155. https://doi.org/10.3354/ab00145

Bik, H. M. (2019). Microbial metazoa are microbes too. *Msystems*, *4*(3), e00109-19. https://doi.org/10.1128/mSystems.00109-19

Bik, H. M., Fournier, D., Sung, W., Bergeron, R. D., & Thomas, W. K. (2013). Intra-genomic variation in the ribosomal repeats of nematodes. *PLoS ONE*, *8*(10), e78230. https://doi.org/10.1371/journal.pone.0078230

Bik, H. M., Lambshead, P. J. D., Thomas, W. K., & Lunt, D. H. (2010). Moving towards a complete molecular framework of the Nematoda: A focus on the Enoplida and early-branching clades. *BMC Evolutionary Biology*, *10*, 353. https://doi.org/10.1186/1471-2148-10-353

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, *27*(4), 233–243. https://doi.org/10.1016/j.tree.2011.11.010

Bik, H. M., Sung, W., De Ley, P., Baldwin, J. G., Sharma, J., Rocha-Olivares, A., & Thomas, W. K. (2012). Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Molecular Ecology*, *21*(5), 1048–1059. https://doi.org/10.1111/j.1365-294X.2011.05297.x

Blaxter, M. (2016). Imagining Sisyphus happy: DNA barcoding and the unnamed majority. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702), 20150329. https://doi.org/10.1098/rstb.2015.0329

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., … Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature*, *392*(6671), 71–75.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1935–1943. https://doi.org/10.1098/rstb.2005.1725

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bowman, J. S., & Ducklow, H. W. (2015). Microbial communities can be described by metabolic structure: A general framework and application to a seasonally variable, depth-stratified microbial community from the coastal west Antarctic peninsula. *PLoS ONE*, *10*, e0135868. https://doi.org/10.1371/journal.pone.0135868

Brabec, J., Kuchta, R., Scholz, T., & Littlewood, D. T. J. (2016). Paralogues of nuclear ribosomal genes conceal phylogenetic signals within the invasive Asian fish tapeworm lineage: Evidence from next generation sequencing data. *International Journal for Parasitology*, *46*(9), 555–562. https://doi.org/10.1016/j.ijpara.2016.03.009

Buckler, E. S. 4th, Ippolito, A., & Holtsford, T. P. (1997). The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics*, *145*(3), 821–832.

Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., … Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, *8*(17), 8908–8920. https://doi.org/10.1002/ece3.4283

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., … Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, *47*, e103. https://doi.org/10.1093/nar/gkz569

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Clarke, K. R., & Gorley, R. N. (2015). *Getting started with PRIMER v7. PRIMER-E*. Plymouth, UK: Plymouth Marine Laboratory.

Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-davis, R. M., Sung, W., Power, D. M., … Thomas, W. K. (2010). Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology*, *19*(Suppl 1), 4–20. https://doi.org/10.1111/j.1365-294X.2009.04473.x

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

De Ley, P., De Ley, I. T., Morris, K., Abebe, E., Mundo-Ocampo, M., Yoder, M., … Thomas, W. K. (2005). An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1945–1958. https://doi.org/10.1098/rstb.2005.1726

Decraemer, W., & Rho, H. S. (2013). Handbook of Zoology. In A. Schmidt-Rhaesa (Ed.), *Order Desmoscolecida* (Vol. 2: Nematoda, pp. 351–372). Berlin: De Gruyter. https://www.degruyter.com/view/books/9783110274257/9783110274257.351/9783110274257.351.xml

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.

Dell'Anno, A., Carugati, L., Corinaldesi, C., Riccioni, G., & Danovaro, R. (2015). Unveiling the biodiversity of deep-Sea nematodes through metabarcoding: Are we ready to bypass the classical taxonomy? *PLoS ONE*, *10*(12), e0144928. https://doi.org/10.1371/journal.pone.0144928

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Eickbush, T. H., & Eickbush, D. G. (2007). Finely orchestrated movements: Evolution of the ribosomal RNA genes. *Genetics*, *175*(2), 477–485. https://doi.org/10.1534/genetics.107.071399

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644. https://doi.org/10.7717/peerj.4644

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., … Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, *1*, 98. https://doi.org/10.1038/ncomms1095

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., & de Groot, G. A. (2015). Not all are free-living: High-throughput DNA

metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Molecular Ecology*, *24*, 4556–4569. https://doi.org/10.1111/mec.13238

Geisen, S., Snoek, L. B., ten Hooven, F. C., Duyts, H., Kostenko, O., Bloem, J., … van der Putten, W. H. (2018). Integrating quantitative morphological and qualitative molecular methods to analyse soil nematode community responses to plant range expansion. *Methods in Ecology and Evolution*, *9*(6), 1366–1378. https://doi.org/10.1111/2041-210X.12999

Gerlach, S. A. (1977). Attraction to decaying organisms as a possible cause for patchy distribution of nematodes in a Bermuda beach. *Ophelia*, *16*(2), 151–165. https://doi.org/10.1080/00785326.1977.10425467

Gibson, A. K., & Morran, L. T. (2017). A model for evolutionary ecology of disease: The case for *Caenorhabditis* nematodes and their natural parasites. *Journal of Nematology*, *49*(4), 357–372. https://doi.org/10.21307/jofnem-2017-083

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., … Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences USA*, *111*(22), 8007–8012. https://doi.org/10.1073/pnas.1406468111

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(Database issue), D597–D604.

Handley, L. L. (2015). How will the "molecular revolution" contribute to biological recording? *Biological Journal of the Linnean Society*, *115*, 750–766. https://doi.org/10.1111/bij.12516

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., … Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, *25*(13), 3101–3119. https://doi.org/10.1111/mec.13660

Holovachov, O., Haenel, Q., Bourlat, S. J., & Jondelius, U. (2017). Taxonomy assignment approach determines the efficiency of identification of OTUs in marine nematodes. *Royal Society Open Science*, *4*, 170315. https://doi.org/10.1098/rsos.170315

Ibal, J. C., Pham, H. Q., Park, C. E., & Shin, J.-H. (2019). Information about variations in multiple copies of bacterial 16S rRNA genes may aid in species identification. *PLoS ONE*, *14*(2), e0212090. https://doi.org/10.1371/journal.pone.0212090

Jensen, P. (1987). Feeding ecology of free-living aquatic nematodes. *Marine Ecology Progress Series*, *35*(1/2), 187–196. https://doi.org/10.3354/meps035187

Kanzaki, N., Giblin-Davis, R. M., Scheffrahn, R. H., Taki, H., Esquivel, A., Davies, K. A., & Herre, E. A. (2012). Reverse taxonomy for elucidating diversity of insect-associated nematodes: A case study with termites. *PLoS ONE*, *7*(8), e43865. https://doi.org/10.1371/journal.pone.0043865

Keller, I., Veltsos, P., & Nichols, R. A. (2008). The frequency of rDNA variants within individuals provides evidence of population history and gene flow across a grasshopper hybrid zone. *Evolution*, *62*(4), 833–844. https://doi.org/10.1111/j.1558-5646.2008.00320.x

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Kumar, V., Dickey, A. M., Seal, D. R., Shatters, R. G., Osborne, L. S., & McKenzie, C. L. (2017). Unexpected high intragenomic variation in two of three major pest thrips species does not affect ribosomal internal transcribed spacer 2 (ITS2) utility for thrips identification. *International Journal of Molecular Sciences*, *18*(10), 2100. https://doi.org/10.3390/ijms18102100

Lallias, D., Hiddink, J. G., Fonseca, V. G., Gaspar, J. M., Sung, W., Neill, S. P., … Creer, S. (2015). Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME Journal*, *9*(5), 1208–1221. https://doi.org/10.1038/ismej.2014.213

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., … Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. https://doi.org/10.1038/nbt.2676

Lanner, J., Curto, M., Pachinger, B., Neumüller, U., & Meimberg, H. (2019). Illumina midi-barcodes: Quality proof and applications. *Mitochondrial DNA Part A*, *DNA Mapping*, *Sequencing*, *and Analysis*, *30*(3), 490–499. https://doi.org/10.1080/24701394.2018.1551386

Leasi, F., Sevigny, J. L., Laflamme, E. M., Artois, T., Curini-Galletti, M., de Jesus Navarrete, A., … Thomas, W. K. (2018). Biodiversity estimates and ecological interpretations of meiofaunal communities are biased by the taxonomic approach. *Communications Biology*, *1*, 237. https://doi.org/10.1038/s42003-018-0119-2

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T. A., Black, K. D., & Pawlowski, J. (2015). High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports*, *5*, 13932. https://doi.org/10.1038/srep13932

Lindner, D. L., & Banik, M. T. (2011). Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia*, *103*(4), 731–740.

Lindner, D. L., Carlsen, T., Henrik Nilsson, R., Davey, M., Schumacher, T., & Kauserud, H. (2013). Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, *3*(6), 1751–1764. https://doi.org/10.1002/ece3.586

Macheriotou, L., Guilini, K., Bezerra, T. N., Tytgat, B., Nguyen, D. T., Phuong Nguyen, T. X., … Derycke, S. (2019). Metabarcoding free-living marine nematodes using curated 18S and CO1 reference sequence databases for species-level taxonomic assignments. *Ecology and Evolution*, *9*, 1211–1226. https://doi.org/10.1002/ece3.4814

Markmann, M., & Tautz, D. (2005). Reverse taxonomy: An approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Philosophical Transactions of the Royal Society of London. Series B*, *Biological Sciences*, *360*(1462), 1917–1924. https://doi.org/10.1098/rstb.2005.1723

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Miljutin, D. M., Gad, G., Miljutina, M. M., Mokievsky, V. O., Fonseca-Genevois, V., & Esteves, A. M. (2010). The state of knowledge on deep-sea nematode taxonomy: How many valid species are known down there? *Marine Biodiversity*, *40*(3), 143–159. https://doi.org/10.1007/s12526-010-0041-4

Moens, T., Bouillon, S., & Gallucci, F. (2005). Dual stable isotope abundances unravel trophic position of estuarine nematodes. *Journal of the Marine Biological Association of the United Kingdom*, *85*(6), 1401–1407. https://doi.org/10.1017/S0025315405012580

Moens, T., & Vincx, M. (1997). Observations on the feeding ecology of estuarine nematodes. *Journal of the Marine Biological Association of the United Kingdom*, *77*(1), 211–227. https://doi.org/10.1017/S0025315400033889

Nieto Feliner, G., Gutiérrez Larena, B., & Fuertes Aguilar, J. (2004). Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). *Annals of Botany*, *93*(2), 189–200. https://doi.org/10.1093/aob/mch027

Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., … Lamberti, G. A. (2016). Estimating species richness using

environmental DNA. *Ecology and Evolution*, *6*(12), 4214–4226. https://doi.org/10.1002/ece3.2186

Orgiazzi, A., Dunbar, M. B., Panagos, P., de Groot, G. A., & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: Opportunities and challenges. *Soil Biology & Biochemistry*, *80*, 244–250. https://doi.org/10.1016/j.soilbio.2014.10.014

Pawlowski, J., Lejzerowicz, F., & Esling, P. (2014). Next-generation environmental diversity surveys of foraminifera: Preparing the future. *Biological Bulletin*, *227*(2), 93–106. https://doi.org/10.1086/BBLv227n2p93

Pereira, T. J., & Baldwin, J. G. (2016). Contrasting evolutionary patterns of 28S and ITS rRNA genes reveal high intragenomic variation in *Cephalenchus* (Nematoda): Implications for species delimitation. *Molecular Phylogenetics and Evolution*, *98*, 244–260. https://doi.org/10.1016/j.ympev.2016.02.016

Platt, H. M., & Warwick, R. M. (1988). *Freeliving marine nematodes: Pictorial key to world genera and notes for the identification of British species*. *British chromadorids*. Leiden: Brill/Backhuys for the Linnean Society of London and the Estuarine and Brackish-Water Sciences Association.

Pochon, X., Wecker, P., Stat, M., Berteaux-Lecellier, V., & Lecellier, G. (2019). Towards an in-depth characterization of Symbiodiniaceae in tropical giant clams via metabarcoding of pooled multi-gene amplicons. *PeerJ*, *7*, e6898. https://doi.org/10.7717/peerj.6898

Porazinska, D. L., Giblin-Davis, R. M., Esquivel, A., Powers, T. O., Sung, W., & Thomas, W. K. (2010). Ecometagenetics confirm high tropical rainforest nematode diversity. *Molecular Ecology*, *19*(24), 5521–5530. https://doi.org/10.1111/j.1365-294X.2010.04891.x

Porazinska, D. L., Giblin-davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., … Thomas, W. K. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, *9*(6), 1439–1450. https://doi.org/10.1111/j.1755-0998.2009.02611.x

Porazinska, D. L., Giblin-Davis, R. M., Sung, W., & Thomas, W. K. (2010). Linking operational clustered taxonomic units (OCTUs) from parallel ultra sequencing (PUS) to nematode species. *Zootaxa*, *2427*(1), 55–63. https://doi.org/10.11646/zootaxa.2427.1.6

Qiao, K., Bai, M., Wang, Q., Hou, X., Chen, J., Xiao, S., … Qing, X. (2019). Unexpected rDNA divergence between two morphologically minimalistic nematodes with description of a new species (Tylenchomorpha: Tylenchidae). *Nematology*, *21*(1), 57–70. https://doi.org/10.1163/15685411-00003195

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596.

R Core Team. (2019). *R: A language and environment for statistical computing (version 3.6.0)*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Sapir, A., Dillman, A. R., Connon, S. A., Grupe, B. M., Ingels, J., Mundo-Ocampo, M., … Sternberg, P. W. (2014). Microsporidia-nematode associations in methane seeps reveal basal fungal parasitism in the deep sea. *Frontiers in Microbiology*, *5*, 43. https://doi.org/10.3389/fmicb.2014.00043

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schuelke, T., Pereira, T. J., Hardy, S. M., & Bik, H. M. (2018). Nematode-associated microbial taxa do not correlate with host phylogeny, geographic region or feeding morphology in marine sediment habitats. *Molecular Ecology*, *27*(8), 1930–1951.

Shaw, J. L. A., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, *197*, 131–138. https://doi.org/10.1016/j.biocon.2016.03.010

Smith, K. F., Kohli, G. S., Murray, S. A., & Rhodes, L. L. (2017). Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *New Zealand Journal of Marine and Freshwater Research*, *51*(4), 555–576. https://doi.org/10.1080/00288330.2017.1298632

Sow, A., Brévault, T., Benoit, L., Chapuis, M.-P., Galan, M., Coeur d'acier, A., … Haran, J. (2019). Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports*, *9*(1), 3646. https://doi.org/10.1038/s41598-019-40243-z

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., … Waterston, R. H. (2003). The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biology*, *1*(2), E45. https://doi.org/10.1371/journal.pbio.0000045

Sun, D.-L., Jiang, X., Wu, Q. L., & Zhou, N.-Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Applied and Environmental Microbiology*, *79*(19), 5962–5969. https://doi.org/10.1128/AEM.01282-13

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Toju, H., & Baba, Y. G. (2018). DNA metabarcoding of spiders, insects, and springtails for exploring potential linkage between above- and below-ground food webs. *Zoological Letters*, *4*, 4. https://doi.org/10.1186/s40851-018-0088-9

Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, *24*(Part 3), 185–196. https://doi.org/10.1016/j.margen.2015.08.007

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Berlin, Germany: Springer.

Wieser, W. (1953). Die Beziehungen zwichen Mundhohlengestalt, Ernahrungsweise und Vorkommen bei freilebenden marinen Nematoden. *Arkiv För Zoologi* (Ser 2), *4*, 439–484.

Yu, G., Lam, T.-T.-Y., Zhu, H., & Guan, Y. (2018). Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Molecular Biology and Evolution*, *35*(12), 3041–3043.

Zar, J. H. (2010). *Biostatistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pereira TJ, De Santiago A, Schuelke T, Hardy SM, Bik HM. The impact of intragenomic rRNA variation on metabarcodingderived diversity estimates: A case study from marine nematodes. *Environmental DNA*. 2020;00e1–16. https://doi.org/10.1002/edn3.77