Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth

Janina Schenk1, Stefan Geisen2, Nils Kleinbölting3, Walter Traunspurger1

1 Department of Animal Ecology, Bielefeld University, Konsequenz 45, 33615, Bielefeld, Germany
2 Laboratory of Nematology, Wageningen University and Research, 6700 ES Wageningen/ Department of Terrestrial Ecology, Netherlands Institute of Ecology NIOO-KNAW, 6708, PB Wageningen, The Netherlands
3 Center for Biotechnology, Bielefeld University, Universitaetsstrasse 25, 33615, Bielefeld, Germany

Corresponding author: Janina Schenk (janina.schenk@uni-bielefeld.de)

Abstract

Microscopic organisms are the dominant and most diverse organisms on Earth. Nematodes, as part of this microscopic diversity, are by far the most abundant animals and their diversity is equally high. Molecular metabarcoding is often applied to study the diversity of microorganisms, but has yet to become the standard to determine nematode communities. As such, the information metabarcoding provides, such as in terms of species coverage, taxonomic resolution and especially if sequence reads can be linked to the abundance or biomass of nematodes in a sample, has yet to be determined. Here, we applied metabarcoding using three primer sets located within ribosomal rRNA gene regions to target assembled mock-communities consisting of 18 different nematode species that we established in 9 different compositions. We determined abundances and biomass of all species added to examine if relative sequence abundance or biomass can be linked to relative sequence reads. We found that nematode communities are not equally represented by the three different primer sets and we found that relative read abundances almost perfectly correlated positively with relative species biomass for two of the primer sets. This strong biomass-read number correlation suggests that metabarcoding reads can reveal biomass information even amongst more complex nematode communities as present in the environment and possibly can be transferred to better study other groups of organisms. This biomass-read link is of particular importance for more reliably assessing nutrient flow through food-webs, as well as adjusting biogeochemical models through user-friendly and easily obtainable metabarcoding data.

Key Words

Nematodes; quantification; community analysis; Illumina MiSeq; species diversity

Introduction

Over the last centuries, researchers aimed at capturing the planet’s biodiversity that consists of about 1.9 million described species – a fraction compared to the diversity of undescribed species (Hajibabaei et al. 2011). Soil and sediments host a huge part of biodiversity and organism biomass on earth (Bar-On et al. 2018). Yet, much of this biodiversity, especially of microorganisms (e.g. bacteria, protists) and microscopic animals (e.g. rotifers, nematodes) remains undescribed and estimates on their numbers and biomass vary profoundly (Fierer et al. 2017; Bar-On et al. 2018). This is due to the fact that their identification is difficult, time-consuming and requires expert-knowledge (Coissac et al. 2012; Jörger et al. 2012). Amongst microscropic animals, nematodes are by far the most abundant (Traunspurger et al. 2012; van den Hoogen et al. 2019). In addition, the functional role of nematodes is diverse as they provide key ecosystem functioning, such as an important link in the food web between bacteria and larger organisms, as well as nutrient cycling (Rutger et al. 1995; Traunspurger et al. 1997; Schmid and Schmid-Araya 2002).
Nematode communities (and those resulting from nematode abundance and diversity estimates) are characterised by morphological taxon identification using microscopy. This is in contrast to comparable studies focusing on microorganisms, which are entirely based on molecular approaches, particularly metabarcoding (Knight et al. 2018; Nilsson et al. 2019). Assessing the community structure of most animal groups, including nematodes using metabarcoding, is promising but not fully developed (Porazinska et al. 2009; Darby et al. 2013; Holovachov et al. 2017; Griffiths et al. 2018; Pačo et al. 2018; Teonis et al. 2018; Weigand and Macher 2019). Differences in copy-numbers of targeted genes and differential taxon amplification, due to imperfect primer matches and variation in amplicon sizes between targeted taxa, are amongst the reasons that artificially change the observed community composition when metabarcoding is used (Prokopowich et al. 2003; Kembel et al. 2012; Bik et al. 2013; Darby et al. 2013). Taxon-assignments also suffer from inaccurate or incomplete reference databases (Cowart et al. 2015; Piňol et al. 2015; Leese et al. 2016). Amongst the only marker genes that are well covered for most microorganisms and nematodes are ribosomal genes (Peham et al. 2017). Currently, 27,287 18S and 21,362 28S rRNA gene reads are available for nematodes in NCBI GenBank (Benson et al. 2013; Nov 2019).

Several studies investigated the performance of metabarcoding for community analyses, but mostly focusing on samples without an a priori knowledge of species composition initially present in the sample as reported for ciliates and microbial communities (Logares et al. 2014; Dong et al. 2017; Pitsch et al. 2019). Studies testing metabarcoding on nematode mock communities—communities with a known composition, often used a limited number of species (n < 10) or a single primer-pair for amplification (Porazinska et al. 2010; Darby et al. 2013; Macherirotou et al. 2019; Waeyenberge et al. 2019). These studies identified a potential link between sequence number and nematode abundance, but no satisfying consensus was found. It remains to be determined if other parameters, such as biomass, might be better linked to read abundance as shown for copepods (Hirai et al. 2015; Clarke et al. 2017), while others failed to find this link (Harvey et al. 2017). Most of the studies listed above were based on family-level, while an accurate inspection at higher taxonomic resolution is missing. A reliable approach to infer biomass data is, however, needed to incorporate into food-web models and implement in elemental flow measurements (Bittleston et al. 2015; Clarke et al. 2017). A positive link of relative read abundance with biomass, with larger, normally less abundant nematodes having amongst the highest read numbers, was recently suggested in a pan-European field survey of soil nematode communities using a 18S rDNA gene region (Wilschut et al. 2019).

In this study, we used mock communities, consisting of 18 different nematode species and applied Illumina MiSeq sequencing, targeting three commonly used barcoding regions within the rRNA gene region. We changed the composition of those species, such as by adding them in equal abundances, by compositions as found in nature and by replacing larger adult with smaller juvenile specimens to assess and calibrate abundance depiction efficiency. We hypothesised (1) that relative taxon composition and revealed diversity depend on the primer pair, due to differences in resolution and PCR-induced differential amplification, especially for communities. Additionally, we hypothesised (2) that relative read abundance is best reflected by relative biomass rather than abundance data due to the differences in ribosomal copy number variation and the resulting increases in barcoding gene numbers.

Material and methods

Nematode culturing and community design

Nematode species were raised on cultured agar plates (1.7%) dosed with cholesterol and E. coli and synchronised regarding life stages. In total, 18 species were used and organisms were individually removed with needles from plates and rinsed in a water drop in order to remove excess agar parts, bacteria and fungi. Organisms were then transferred into an Eppendorf tube containing lysis buffer (Machery & Nagel, Nucleo Spin Tissue XS Kit), according to predefined mock-community ratios (Table 1). Each tube contained 198–200 nematodes from the 18 species, resulting in 9 different mock-communities with 3 replicates for each mock-community. Mock-communities represented ratios that are found in nature (Peters and Traunspurger 2005; Kazemi-Dinan et al. 2014), as well as evenly distributed ratios in order to examine sequencing success. Subsequently, 8 µl of proteinase-K was added and samples were lysed at 56 °C within a rocking water bath for 8 hours. Subsequently, DNA was extracted, according to the NucleoSpin Tissue Kit XS protocol (Macherey & Nagel, Hilden). Prior to amplification, DNA concentration was checked with PicoGreen. Samples were amplified with three different primer pairs that had been commonly used for metabarcoding of microscopic animals previously. The 18S rRNA gene was amplified with F04/R22 (5’-GCTTGCTCTCAAGATTA-AGCC-3’/5’-GCTTGCTGCTTTCTTGGGA-3’) (Fonseca et al. 2010), amplifying a ~350 bp fragment of the V1-V2 region and 3NDF/C_1132F (5’-GGCAAGTCTGGTGC- CAG 3’/5’-TCCGTCAAATTTCTTAAATG 3’), amplifying a ~530 bp fragment of the V4 region (Geisen et al. 2018). A ~520 bp fragment of the D3-D5 region of the 28S rRNA gene was amplified with the primer pair 1274/706 (5’-GACCCGTCTTGAACACCGGA-3’/5’-GCACGTTCTGCCACGCA-3’) introduced by Markmann and Tautz (2005), which has proven to work extremely well for freshwater nematodes (Ristau et al. 2013; Schenk et al. 2016). Primers pairs are subsequently simplified to...
Table 1. Mock-community composition of 18 nematode species with a total of 198–200 individuals present in each assembled community. Nine different communities were created, all with at least one of each species present. One species (Rhomborhabditis regina) was added either as adults or juveniles, while only adults of all other species were added. Furthermore, the wet weight (ww) of the 18 species is given in µg, together with the range of biomass. Each sample was replicated 3 times, resulting in 27 mock-communities.

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of individuals</th>
<th>Sample</th>
<th>Biomass ww</th>
<th>Biomass range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhomborhabditis regina adult (Schulte &amp; Poinar, 1991)</td>
<td>11 0 6 2 5 10 5 10 2 20.29 18.82–34.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhomborhabditis regina juvenile</td>
<td>0 11 6 0 0 0 0 0 0 6.09 3.25–16.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caenorhabditis elegans Maupas 1900</td>
<td>11 11 11 166 41 15 115 70 21 3.74 2.22–5.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panagrellus redivivus (Goody, 1943)</td>
<td>11 11 11 2 4 10 5 10 23 3.83 3.44–5.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pristionchus pacificus Sommer et al. 1996</td>
<td>11 11 11 2 5 10 5 10 13 3.10 2.62–3.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pristionchus entomophaga (Steiner, 1929)</td>
<td>11 11 11 2 15 15 10 5 7 4.02 3.45–4.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plectus cf. acuminatus Bastian, 1865</td>
<td>11 11 11 2 3 10 5 5 3 1.98 1.26–3.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plectus aquatilis Andraossy, 1985</td>
<td>11 11 11 2 5 5 5 2 1.50 1.15–2.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paragolaimella bernenstei (Steiner, 1914)</td>
<td>11 11 11 2 20 15 5 5 32 1.37 1.16–1.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panagrolaimus thienemanni Hirschmann, 1952</td>
<td>11 11 11 2 5 5 5 10 10 0.31 0.28–0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrostichus sp.</td>
<td>11 11 11 2 5 10 5 5 2 0.79 0.45–1.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrostichus nudicapitatus (Steiner, 1914)</td>
<td>11 11 11 2 5 25 5 5 2 0.92 0.68–1.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poikilolaimus regensfasi Sudhaus, 1980</td>
<td>11 11 11 2 15 5 5 10 11 0.65 0.44–0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poikilolaimus oxycerca (de Man, 1895)</td>
<td>11 11 11 2 5 5 5 5 5 2 0.73 0.70–0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrobeiloides tricornus (Thorne, 1925)</td>
<td>11 11 11 2 25 10 5 10 9 0.82 0.73–1.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrobeiloides cf. nanus (de Man,1880)</td>
<td>11 11 11 2 10 15 5 10 23 0.77 0.62–1.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diploscapter coronatus (Cobb, 1893)</td>
<td>11 11 11 2 5 15 5 5 2 0.24 0.20–0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aphelenchoides parietinus Steiner, 1932</td>
<td>11 11 11 2 25 10 5 10 34 0.14 0.09–0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sum                                          | 198 198 199 200 200 200 200 200 200 |

18S_V1 (F04/R22), 18S_V4 (3NDf/C_1132f) and 28S_D3 (1274/706). PCR conditions were 30 cycles with 60 s pre-denaturation at 96 °C, followed by 15 s at 96 °C, 30 s at 58 °C and 90 s at 70 °C. In a second PCR with the same conditions, Illumina indices were attached in 10 cycles. Metabarcoding was carried out at LGC Genomics (Berlin) on an Illumina MiSeq (V3 2 × 300 bp). About 20 ng of each sample was pooled for sequencing. Samples were delivered demultiplexed.

Reference sequences

Individual species were amplified, Sanger-sequenced and used as reference in the bioinformatic pipeline (see below). PCR conditions followed Schenk et al. (2016). Sequences are deposited at NCBI GenBank under the Accession numbers: MK543220–MK543233, MK541669–MK541684, MK541653–MK541668. The closely related species Acrobeiloides tricornus and Acrobeiloides cf. nanus, Plectus aquatilis and Plectus cf. acuminatus and Plectus velox, as well as Acrostichus nudicapitatus and Acrostichus sp. could not be distinguished, based on their sequences for the 18S rRNA gene region. For the 28S rDNA gene region, Plectus aquatilis and Plectus cf. acuminatus, as well as Acrostichus nudicapitatus and Acrostichus sp. could not be distinguished. However, as natural communities contain closely related species as well, this was also maintained in our study. For two species, Panagrolaimus thienemanni and Aphelenchoides parietinus, no Sanger reference sequence could be generated for any of the markers (Suppl. material 1: Table S1). For the latter, NCBI sequences were available for the 28S rDNA gene region (MF325173.1, MF325174.1)

Bodyweight correlation

Mean biomass (wet weight) of every species was calculated following (Andrassy 1956) for several individuals (n = 10–15 adults) that were heat-fixated and measured (Table 1). Furthermore, the proportion of generated reads for 18S_V4 and 28S_D3 was plotted against the relative biomass and the relative abundance together with linear regression in MatLab (MATLAB User’s Guide 1998). The correlation for absolute biomass and absolute abundance against the read numbers are given in Suppl. material 2: Fig. S1.

Bioinformatic analysis

Except for the taxonomic classification and primer removal, the MiSeq standard operation procedure using mothur (Kozich et al. 2013; Schloss et al. 2009) was used. Demultiplexed reads were combined to reach a longer paired-end read for higher phylogenetic resolution by using the make.contigs function of mothur with default settings, thereby correcting sequencing errors in the overlapping region. Primer sequences were then removed from the combined reads using cutadapt with a default error rate of 0.1 (Martin 2011). Reads with ambiguous bases, homopolymers larger than 10 bases and of unexpected short or long length (allowed range: 333–367 for 18S_V1, 514–597 for 18S_V4 and 471–516 for
Community structure & Statistical community analysis

For each primer pair, read numbers for each species in each replicate were averaged to decrease variation and communities were compared to relative abundance and relative biomass, based on the number of inoculated specimens. For statistical analyses of communities, Bray Curtis similarity was applied with non-transformed data. Non-metric multidimensional scaling (NMDS) plots were created with MatLab (MATLAB User’s Guide 1998), while Permutational multivariate analysis of variance (PERMANOVA) was conducted with PRIMER-E v6 (Clarke and Gorley 2006). PERMANOVA was computed with 9,999 permutations and pairwise comparisons amongst relative abundance, biomass and the read proportions for the three primer pairs. The samples were used as the nested factor in the treatments (e.g. abundance, biomass or primer used). A list of all pairwise comparisons is given in Suppl. material 1: Table S3. The effect of amplicon length and mismatches on sequencing success was analysed with Analysis of Variance (ANOVA) in R (R Core Team 2013) in order to check if there was an significant effect by the two factors on the sequencing success. Sequencing success was defined as expected to encountered sequencing ratios based on biomass proportions.

Results

Amplification success

The three primer pairs had a different amplification success (Table 2) with the 28S_D3 marker resulting in the highest recovery rate of species present in the mock-communities (n = 14), followed by 18S_V4 (n = 12), while the 18S_V1 returned the lowest OTU number (n = 10). Different amplification success could partly be attributed to mismatches in the primer regions, such as for Diploopus coronatus, which could not be amplified within the metabarcodeing approach for any of the markers, although it could be amplified with Sanger-sequencing (Suppl. material 1: Tables S1, S4). ANOVA showed that variations in length of the amplified region of different nematode taxa significantly affected amplification success of 18S_V4 (p < 0.001), the primer pair that differed most in length between nematode taxa (shortest amplicon of 492 bp in 531 bp). For the other two markers, amplicon length (28S_D3: p = 0.41, 18S_V4: p = 0.21) and mismatches (28S_D3: p = 0.51, 18S_V4: p = 0.6) did not significantly affect amplification success (28S_D3: 512 bp – 525 bp, 18S_V1: 333 bp – 364 bp, Suppl. material 1: Table S5). We focused our analyses on all sequences assigned to the nematode taxa that were added, as contamination with non-target sequences was low (0.5–2.1%, Table 2).

Community structure

The community structure differed between the primer pairs, with 28S_D3 resulting in the most similar community depiction to the initial inoculated communities, while depiction for the 18S_V1 primer pair substantially differed. Nematode communities amplified with the 28S_D3 primer pair grouped together with nematode mock communities presented as biomass data in the NMDS plot (Fig. 1a), which was also supported by pairwise comparisons between 28S_D3 and biomass showing similar patterns (t-value = 0.8193, p = 0.5293). In contrast, nematode abundances differed from the community composition as depicted by the primer pair 28S_D3 (Fig. 1a) and supported statistically (t-value = 2.2339, p = 0.0076). The communities for the 18S_V4 primer pair were different from mock-community data presented as biomass (t-value = 1.9353, p = 0.0279) and even more to abundance (t-value = 3.6898, p < 0.001, Fig.
1b). Communities revealed with the primer pair 18S_V1 were most different to biomass and abundance (t-value = 5.9545, p < 0.001 and 6.7753, p < 0.001, Fig. 1c) and were therefore not evaluated further. The community structure of the inoculated communities differed when represented as biomass or abundance (t-value = 2.281, p = 0.0068). Primer pairs 28S_D3 and 18S_V4 revealed nematode communities in a similar way (t-value = 1.278, p = 0.1638). In contrast, all communities amplified with the primer pair 18S_V1 were strikingly different, being characterised by a clear dominance of few species, for example, *Rhomborhabditis regina* and *Plectus aquatilis* (Suppl. material 2: Fig. S3).

**Table 2.** Species recovered by the three primer pairs. Given are the species added in the mock-communities and the performance of each marker. An “x” indicates that the species was found, an “n” means the species was missing and an (x) shows that the nucleotide sequence of this species was identical to another species, indicated in bold and, therefore, could not be experimentally verified at the species level in the metabarcoding approach. Furthermore, the total number of OTUs that could be recovered from the mock community species, including species with identical sequences, as well as without genetically indistinguishable species (in parentheses), are given, together with the average read length (bp) and the contamination (in %).

<table>
<thead>
<tr>
<th>Species</th>
<th>18S_V1</th>
<th>18S_V4</th>
<th>28S_D3</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Rhomborhabditis regina</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Panagrellus redivivus</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Pristionchus pacificus</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Pristionchus entomophagus</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Aphelenchoides parietinus</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Acrobeloides cf. nanus</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Acrobeloides tricornus</em></td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
</tr>
<tr>
<td><em>Panagrolaimus thienemanni</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Poikilolaimus cf. regenfussi</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Poikilolaimus oxyacera</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Paroigolaimella bernensis</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Acrostichus sp.</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Acrostichus nudicapitatus</em></td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td><em>Diploscapter coronatus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Plectus aquatilis</em></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Plectus cf. acuminatus</em></td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td><em>Plectus velox</em></td>
<td>(x)</td>
<td>(x)</td>
<td>x</td>
</tr>
<tr>
<td>Number of OTUs</td>
<td>13(10)</td>
<td>16 (12)</td>
<td>16 (14)</td>
</tr>
<tr>
<td>Avg. read length (bp)</td>
<td>405</td>
<td>583</td>
<td>532</td>
</tr>
<tr>
<td>Avg. contamination (%)</td>
<td>0.53</td>
<td>0.81</td>
<td>2.14</td>
</tr>
</tbody>
</table>

Figure 1. NMDS plots for a) the 28S_D3 region, b) the 18S_V4 region and c) the 19S_V1 region depicting the 9 samples. Each sample is shown with three replicates, together with the relative proportion of biomass (black) and the relative abundance of the original community (cyan).

Quantification

Relative biomass positively correlated with relative read numbers for the primer pair 28S_D3 (R² = 0.90539, p < 0.001) and 18S_V4 (R² = 0.81396, p < 0.001, Fig. 2 a–b). At the level of individual mock-communities, especially samples with dominant species were often positively correlated with relative biomass proportions for the 28S_D3 and 18S_V4 primer pair, for example, Sample 4 (28S_D3: R² = 0.9995, p < 0.001; 18S_V4: R² = 0.9656, p < 0.001). The same was found for Sample 7 (28S_D3: R² = 0.9885, p < 0.001 18S_V4: R² = 0.923, p < 0.001) or Sample 9 (28S_D3: R² = 0.9162, p < 0.001; 18S_V4: R² = 0.3069, p = 0.02597). In general, species with a high contribution to the total biomass resulted in the highest read proportions, while species with a low biomass were clearly less amplified. The biomass of juvenile and adult specimens differed by a factor of 5.00 (Suppl. material 1: Table S6 and Suppl. material 2: Fig. S4), while the amplification of different life stages of *R. regina* also resulted in varying read proportions, with the adult nematodes generating significantly more reads than the juvenile nematodes (Wilcoxon-Mann-Whitney test, W = 81, p < 0.001). For the 28S_D3 primer pair, 17.14 times more reads were generated, while the 18S_V4 primer pair resulted in 4.65 times more reads and 5.81 times.

Relative species abundances were positively correlating with read numbers for the primer pairs 28S_D3 (R² = 0.53694, p < 0.001) and 18S_V4 (R² = 0.33077, p < 0.001), but less strong than for relative biomass-read number correlations (Fig. 2c–d). Correlations of absolute biomass and absolute abundance against absolute read counts showed a similar pattern, but with overall weaker correlation strengths (Suppl. material 2: Fig. S1).
Discussion

In this study, we show according to our hypothesis (1) that taxon composition and revealed diversity depend on the applied primer pairs. As furthermore hypothesised, we could show (2) that the relative sequence abundance strongly correlates with the relative taxon biomass for two of the three primer pairs tested.

Resolution and species recovery

A total of 78% of species diversity present in the samples was recovered for the 28S_D3 primer pair, while, for the other primers, resolution was lower (18S_V1: 55%; 18S_V4: 67%). A higher diversity was recovered considering indistinguishable taxa (28S_D3 and 18S_V4: 89%; 18S_V1: 78%). Those that could not be distinguished shared the same marker gene sequence and, as such, cannot be distinguished, based on sequencing (Table 2). The rRNA genes were recently suggested to be incapable of identifying many nematode taxa to species and genus level resolution was consequently suggested as a reliable lower-resolution alternative (Creer et al. 2016; Sahraean et al. 2017). It should, therefore, be considered in metabarcoding studies of unknown species composition, when the full species diversity is aimed at identification in a sample. However, species that are closely related and, as such, share the same genetic sequence commonly can share similar morphological, physiological and functional traits (Potapov et al. 2019). For example, nematode species have proven to be equally sensitive to pollutants within the same genus (Höss et al. 2017). Therefore, the relative-read abundance links at lower taxonomic resolution, which likely links to functional information, still holds. From a total of 18 species used, one species could not be recovered with any of the primer-pairs and other spe-
cies such as *Aphelenchoides parietinus* were under-represented in relative sequence reads. This can be mainly explained by mismatches present in primer sites. Accordingly, a 10 fold drop in amplification efficiency was reported to be caused by only one nucleotide mismatch in the priming region (Piñol et al. 2015). Due to amplicon length variations, *Panagrellus redivivus* (531 bp) might not be amplified with 18S_V4, as the primer length is at the maximum of the capacity our Illumina approach can sequence (2× 300 bp). This is supported by the fact that *P. redivivus* was found in unmerged forward and reverse reads (Suppl. material 1: Table S7). Furthermore, we demonstrate that longer amplicon sizes negatively influenced amplification success for the 18S_V4 marker (Suppl. material 1: Table S4).

**Community structure**

The community depiction for this study varied due to primer-induced differences. The 18S_V1 marker was, in our case, not suitable for distinguishing communities due to a low species recovery and an extreme over-amplification of mainly two species (*Rhomborhabditis regina* and *Plectus aquatilis*) and is, therefore, not discussed further. In turn and in line with most accurate diversity representations, the 28S_D3 primer pair could reliably distinguish between communities, followed by the 18S_V4 marker. The difference for the V4-marker is likely introduced by higher length-variation in this marker (Suppl. material 1: Table S4). Especially, communities with dominant taxa were depicted accurately and uncertainties were mostly detected from less abundant taxa for both primer pairs. Many metabarcoding studies reported a problem in detecting rare taxa, which can be masked by the more abundant taxa during PCR or sequencing (Evans et al. 2016) or even filtered out during bioinformatics analyses due to too low read abundances (Elbrecht et al. 2018).

**Quantification**

Relative biomass and read proportions for almost all mock-community species at the species level were not different for the 28S_D3 primer pair, suggesting that relative biomass can reliably be depicted by relative read numbers obtained by our nematode metabarcoding approach. In support, species that were missing or amplified at low proportions for the 28S_D3 and 18S_V4 markers in the dataset were those that also had a low overall biomass (*Panagrolaimus thienemanni* and *Diploscapter coronatus*, Table 1). These results imply that upsampling with more taxa and more complex communities might be possible. This study covered many of the extremely ubiquitous Rhabditidae, as well as other members of the nematode taxa as presented by van Megan et al. (2007), but more detailed studies using a wider phylogenetic range of nematode taxa are needed to confirm these findings across a wider species-range within Nematoda. If relative sequencing data actually represents relative biomass data, the information provided by metabarcoding might be implemented in studies investigating food-web structures (Sechi et al. 2018). This relationship, although by far less strong, has been observed for nematodes recently (Wilschut et al. 2019), but was limited to a 18S rDNA gene region. The here observed biomass-read number link was further confirmed by differences in read numbers between juvenile and adult specimen of the same species, such as shown before in fishes (Maruyama et al. 2014). As many ecosystem functions rely on biomass estimations, such as production, our results reveal that metabarcoding can be used to assay biomass distribution of previously hardly studied organisms, such as minute organisms living in soils and sediments. While we here show this potential for nematodes, we are convinced that this approach can be applied and provide biomass estimates also for other groups of organisms. Thorough calibration efforts are now needed to reliably link diversity and (relative) biomass of taxa in a sample.

**Acknowledgements**

We thank Steffi Gehner for raising nematodes on cultured plates at synchronised life stages and thank Professor Oliver Krüger for proving laboratory space for the molecular work. We also thank the German Federal Institute of Hydrology (BfG) for funding our research. Bioinformatics work was furthermore supported in parts by grants of the German Federal Ministry of Education and Research (BMBF) of the project “Bielefeld-Gießen Center for Microbial Bioinformatics – BiGi” (Grant-number 031A533) within the German Network for Bioinformatics Infrastructure (de.NBI). S. G. was supported by the NWO-VENI grant from the Netherlands Organisation for Scientific Research (016.Veni.181.078). We acknowledge the constructive comments by one anonymous referee and the editor.

**References**


Potapov AM, Scheu S, Tiunov A V (2019) Trophic consistency of supra
Cunha Castro José Dam M, Djigal D, Escuer M, Griffiths BS, Guti
longicaudata
Trescases AM, Unangst SK, Kepler RM, Buyer JS, Cavigelli MA, Mirsky
Traunspurger W, Goede RGM de, Rutgers M, Brussaard L, Mulder C (2018)


Supplementary material 1
Tables S1–S7
Authors: Schenk J, Geisen S, Kleinbölting N, Traunspurger W
Data type: species data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.3.46704.suppl1

Supplementary material 1
Figures S1–S4
Authors: Schenk J, Geisen S, Kleinbölting N, Traunspurger W
Data type: species data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.3.46704.suppl2