MOLECULAR ECOLOGY

Molecular Ecology (2010), **19** (Suppl. 1), 4–20 doi: 10.1111/j.1365-294X.2009.04473.x

Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises

S. CREER,* V. G. FONSECA,* D. L. PORAZINSKA,† R. M. GIBLIN-DAVIS,† W. SUNG,‡ D. M. POWER,§ M. PACKER,¶ G. R. CARVALHO,* M. L. BLAXTER,** P. J. D. LAMBSHEAD†† and W. K. THOMAS‡

*School of Biological Sciences, Environment Centre Wales, Deiniol Road, College of Natural Sciences, Bangor University, Gwynedd LL57 2UW, UK, †Fort Lauderdale Research and Education Center, University of Florida, IFAS, 3205 College Avenue, Fort Lauderdale, FL 33314, USA, ‡Hubbard Center for Genome Studies, University of New Hampshire, 35 Colovos Rd, Durham, NH 03824, USA, §Centre of Marine Sciences, CCMAR–CIMAR Associate Laboratory, University of Algarve, Gambelas, 8005-139 Faro, Portugal, ¶Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK, **Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK, ††School of Ocean and Earth Science, University of Southampton, National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK

Abstract

Biodiversity assessment is the key to understanding the relationship between biodiversity and ecosystem functioning, but there is a well-acknowledged biodiversity identification gap related to eukaryotic meiofaunal organisms. Meiofaunal identification is confounded by the small size of taxa, morphological convergence and intraspecific variation. However, the most important restricting factor in meiofaunal ecological research is the mismatch between diversity and the number of taxonomists that are able to simultaneously identify and catalogue meiofaunal diversity. Accordingly, a molecular operational taxonomic unit (MOTU)-based approach has been advocated for *en mass* meiofaunal biodiversity assessment, but it has been restricted by the lack of throughput afforded by chain termination sequencing. Contemporary pyrosequencing offers a solution to this problem in the form of environmental metagenetic analyses, but this represents a novel field of biodiversity assessment. Here, we provide an overview of meiofaunal metagenetic analyses, ranging from sample preservation and DNA extraction to PCR, sequencing and the bioinformatic interrogation of multiple, independent samples using 454 Roche sequencing platforms. We report two examples of environmental metagenetic nuclear small subunit 18S (nSSU) analyses of marine and tropical rainforest habitats and provide critical appraisals of the level of putative recombinant DNA molecules (chimeras) in metagenetic data sets. Following stringent quality control measures, environmental metagenetic analyses achieve MOTU formation across the eukaryote domain of life at a fraction of the time and cost of traditional approaches. The effectiveness of Roche 454 sequencing brings substantial advantages to studies aiming to elucidate the molecular genetic richness of not only meiofaunal, but also all complex eukaryotic communities.

Keywords: 454 environmental sequencing, meiofaunal and eukaryotic biodiversity, metagenetics, metagenomics

Received 19 June 2009; revision received 19 August 2009; accepted 21 August 2009

Correspondence: Simon Creer, Fax: +44(0)1248 382569;
E-mail: s.creer@bangor.ac.uk
S. Creer and V. G. Fonseca are joint first authors of this work.

Introduction

Robust, quantified biodiversity assessment is the key to deep understanding of the relationship between biodi-

© 2010 Blackwell Publishing Ltd

versity and ecosystem functioning. The effects of major anthropogenic stressors on global ecosystems, including elevated CO_2, pollution, habitat loss and fragmentation, add urgency to this field, demanding an increasing focus on mechanistic and predictive studies. However, investigating the role of biodiversity in maintaining ecosystem function, resilience and recovery (Sutherland *et al.* 2006) can be meaningfully addressed only if biodiversity can first be identified. The identity of macrofaunal and floral communities can be ascertained by teams of trained taxonomists/ecologists with their skills being augmented by globally integrated molecular barcoding approaches (Hebert *et al.* 2003a; Hajibabaei *et al.* 2007). Similarly, recent advances in sequencing power and the molecular identification of microbes are facilitating the more realistic characterization of the phylogenetic affinities, identity (DNA sequences), composition (Sogin *et al.* 2006; Huber *et al.* 2007), dynamics and even functional capacity (Edwards *et al.* 2006; Mou *et al.* 2008) of prokaryotic communities. The application of second-generation sequencing has also been applied to the identification of protist communities in this edition (Medinger *et al.* 2010; Stoeck *et al.* 2010). There remains, however, a well-acknowledged biodiversity identification gap related to eukaryotic meiofaunal organisms (Blaxter 2003; Blaxter & Floyd 2003; Tautz *et al.* 2003; Blaxter *et al.* 2005).

Meiofaunal taxa are a paraphyletic assemblage, grouped on the basis of size (i.e. organisms that pass through a 0.5-mm sieve but are retained on 25–65 μm sieves). Approximately 60% of animal phyla have meiofaunal representatives and meiofaunal Platyhelminthes, Nemertea, Nematoda, Rotifera, Annelida, Arthropoda, Tardigrada, Mollusca and Chordata have taxa that occupy key roles in marine, freshwater and terrestrial habitats (Higgins & Thiel 1988; Giere 2009). Meiofaunal assemblages are dominated by nematodes and are characterized by high abundances (up to 10^8 individuals per 1 m^2) and diversity (up to 60 morphological species per 75 cm^3 of sediment) (Lambshead 2004). Thus, although meiofaunal organisms are conceptually and demonstrably ecologically important (Snelgrove *et al.* 1997; Danovaro *et al.* 2008), current estimates of global species richness remain a matter of conjecture (Lambshead & Boucher 2003). For nematodes, global estimates of species richness range from 100 000 to 1 000 000, but only ~27 000 species have been described (Platt & Warwick 1983; Coomans 2000; Hugot *et al.* 2001), and contemporary studies routinely recover between 30% and 40% of sampled taxa that are new to science (Lambshead & Boucher 2003). Meiofaunal taxon diversity and abundance is so great that effectively studying communities requires a huge investment in resources and labour. The effort expended in assigning only 10% of nematodes to known species was 120-fold that required to successfully assign all vertebrate morphospecies to known taxa (Lawton *et al.* 1998) in tropical forest habitats.

The identification bottleneck associated with meiofaunal taxonomy is confounded by a range of taxonomic hurdles: the small size and fragility of organisms, convergent evolution, morphological conservatism (Derycke *et al.* 2005, 2008; Bhadury *et al.* 2008; Fontaneto *et al.* 2009) and developmental and sexual variation in morphology (Tautz *et al.* 2003; Lambshead 2004; Blaxter *et al.* 2005). Perhaps the most restricting factor in meiofaunal research is the mismatch between the diversity and abundance of multiple phyla occupying a range of ecological niches and habitats and the number of taxonomists that are able to simultaneously identify and catalogue meiofaunal diversity. In order to address this impediment, it has been suggested that *en mass* molecular identification of meiofaunal communities may significantly advance knowledge and progress in meiofaunal research (Blaxter & Floyd 2003; Markmann & Tautz 2005). Although the molecular identification of meiofaunal communities shares similarities with current molecular barcoding (Hebert *et al.* 2003a,b) and microbial phylotype approaches (Kemp & Aller 2004; Shaw *et al.* 2008), there remains a difference in methodology, taxonomic richness and diversity.

Phylotypes, molecular operational taxonomic units and barcoding for the identification of biodiversity

With a molecular barcoding approach, a standardized homologous region of the genome [e.g. the mitochondrial cytochrome oxidase subunit I gene (COI) for animals] is used for species identification, and is linked to a virtual or actual physical molecular voucher specimen (Hebert *et al.* 2003a; Ratnasingham & Hebert 2007). However, when dealing with individuals or communities of microscopic organisms, the whole voucher specimens are usually sacrificed in order to extract genomic DNA (Blaxter *et al.* 2005; De Ley *et al.* 2005). Advances in video capture technology of microscopic organisms (De Ley *et al.* 2005) and individual organismal PCRs (Floyd *et al.* 2002, 2005; De Ley *et al.* 2005; Bhadury *et al.* 2006; Meldal *et al.* 2007) can overcome this problem and forge a link between taxon ecology/morphology and community-based DNA analyses. Such research provides potential for linking taxonomy, phylogeny (Forest *et al.* 2007; Warwick & Somerfield 2008), functional (Petchey & Gaston 2006) and molecular ecology. It also effectively engages and links morphological taxonomists with molecular ecologists, a connection that will be vital for a holistic approach towards ecosystem-based research. However, standard barcoding approaches are

© 2010 Blackwell Publishing Ltd

not appropriate for large-scale environmental analyses mainly because of extensive abundances and putative hyperdiversity of some taxa (e.g. nematodes, Lambshead 2004; Lambshead & Boucher 2003). Further to this, the extent of taxonomic coverage and lack of taxonomic expertise, manpower and resources makes the task of barcoding environmental samples inefficient.

Instead, the proposed identification of operational taxonomic units (OTUs) in eukaryotic metagenetic analyses has more in common with prokaryotic phylotype (Kemp & Aller 2004) delineation than with species identification using standardized barcoding approaches. The term metagenomics is sometimes used to consider the analysis of any environmentally derived genomic DNA (Hugenholtz & Tyson 2008). Here though, we distinguish between metagenetics, the large-scale analysis of taxon richness via the analysis of homologous genes, and metagenomics, the functional analysis of environmentally derived DNA from unculturable organisms (Edwards *et al.* 2006; Rodriguez-Brito *et al.* 2006; Blow 2008; Hugenholtz & Tyson 2008; Mou *et al.* 2008).

Bacterial phylotypes are groups of sequences that are created by subjecting a larger community of sample-derived shotgun sequences to a user-defined base pair cutoff algorithm. In most cases, phylotypes of a particular grouping (e.g. 97% for bacteria, Venter *et al.* 2004; Shaw *et al.* 2008) are used as a proxy for 'species'. Although microbial communities can be orders of magnitude more diverse than micro-eukaryotic communities, the similarities of their intractable community compositions have led to similar approaches in studying eukaryotic protists (Moon-van der Staay *et al.* 2001; Moreira & Lopez-Garcia 2002) and meiofaunal organisms (Floyd *et al.* 2002; Blaxter & Floyd 2003). For meiofaunal organisms, Floyd *et al.* (2002) formally defined the molecular operational taxonomic unit (MOTU) concept whereby sequences derived from individual specimens are defined as belonging to the same MOTU, based on a user-defined cutoff. The term was later extended to community DNA extractions in Blaxter *et al.* (2005). Normally, the MOTUs do not have any formal correlation with published species descriptions. However, correlations can be achieved by *de novo* elucidation of cryptic species (Abebe & Blaxter 2003), bioinformatic sequence comparisons to existing databases (with both molecular and morphological data), further sequencing or future classifications, termed 'reverse taxonomy' (Markmann & Tautz 2005).

Environmental metagenetics

Until recently, most molecular identification was achieved using Sanger chain-termination sequencing (Kemp & Aller 2004; Venter *et al.* 2004). However, there has recently been a rise in the use of ultrasequencing platforms (Margulies *et al.* 2005) for metagenetic identification of microbial phylotypes using homologous gene regions (Sogin *et al.* 2006; Hall 2007; Huber *et al.* 2007) derived from environmental DNA. The recent increases in sequencing throughput represent a significant shift in our ability to disentangle the biotic complexity of ecosystems. From sample collection to data analysis, there are numerous steps, questions and an exponentially large number of hypotheses that could be tested in order to optimally analyse environmental meiofaunal diversity.

Here, we first provide an overview of the relevant focal areas in an attempt to highlight potential approaches and pitfalls in meiofaunal metagenetics. Second, we present independent data sets derived from ultrasequencing experiments of two different ecological communities; the marine benthos and tropical rain forest habitats. By overviewing separate approaches to environmental metagenetics, we aim to illustrate a range of protocols that can be utilized to analyse contrasting, yet hitherto, inaccessible meiofaunal communities on a scale that has previously not been possible. We aim to illustrate the advantages and limitations of ultrasequencing approaches in addressing large-scale identification of complex eukaryotic communities. Furthermore, we introduce a bioinformatic pipeline that can be used to analyse the data, derived from different but closely related genomic regions, in a computationally expedient fashion. The tropical rain forest case study predominantly targeted nematodes, whereas the marine example targeted collective meiofauna (extended to include organisms ranging from 45 to 1000 μm in size). The approaches and data presented here do not test specific hypotheses regarding metagenetic analyses, but are intended to provide a resource that will be useful to researchers wishing to pursue similar research. Although meiofaunal organisms are the primary focus, the general principles are easily transferrable to other eukaryotic as well as prokaryotic taxa.

Methodological overview and rationale

Sample preservation and extraction

Once an ecologically suitable sampling strategy has been designed, an appropriate decision needs to be made regarding sample processing. Given the diverse and dynamic nature of the micro- and meiofauna, it is predicted that after removing a small subsample of the community, a natural progression of ecological interactions will change the population composition. It is therefore important to either preserve or process samples shortly after collection. Some experiments (e.g. those with small sample sizes or local collection

© 2010 Blackwell Publishing Ltd

regimes) may lend themselves to field processing. Others will necessitate sample, and more importantly, DNA preservation. Such decisions are based predominantly on logistics. Moreover, if sampling regimes are extensive and geographically diverse, it is preferable to deal with a large number of small samples, rather than *vice versa*, thus facilitating statistical flexibility in downstream analyses.

Formalin is the preferred fixative for morphological analyses of the meiofauna (Giere 2009), but specimens fixed in formalin yield low-quality and degraded DNA (but see Thomas *et al.* 1997 and Bhadury *et al.* 2005). Conversely, samples fixed in ethanol may yield DNA optimal for downstream molecular manipulations, but inconsistent amplifications have been observed by researchers working on meiofaunal organisms. Furthermore, ethanol-preserved samples can corrupt morphological features due to osmotically driven shrinkage (Bhadury *et al.* 2006). In an attempt to overcome these constraints, samples are often split between formalin and ethanol preservation to yield distinct samples for morphological and molecular genetic analysis respectively. Adopting a split sampling approach, however, not only creates a problem of potentially unequal community composition between samples, but also precludes obtaining both morphological and molecular data from the same individual (Yoder *et al.* 2006). An answer to the preservation issue is the use of a solution of 20% DMSO, 0.25 M disodium EDTA, saturated with NaCl, pH 8.0, recently known as DESS by Yoder *et al.* (2006). Originally proposed for the preservation of avian blood samples (Seutin *et al.* 1991), DESS has yielded PCR-ready DNA from individual nematodes and communities of entire soil/sediment samples for up to 1 year at room temperature. DESS works by inactivating naturally occurring nuclease activities by a combination of a severe osmotic shock, followed by rapid transportation of disodium EDTA and NaCl into tissues enabled by DMSO (Yoder *et al.* 2006). As with all DNA preservation approaches, it is important to optimize the DNA to buffer ratio to achieve effective DNA preservation. Such is particularly the case for wet soil/sediment samples where the inclusion of significant volumes of sample water may dilute either the concentration of ethanol and/or DESS, preventing complete inhibition of nuclease activities.

Meiofaunal organisms must always be extracted from the substrate because the biomass is orders of magnitude lower than the actual sample volume of soil, sediment or water. Separation can be achieved by employing several approaches (reviewed in Somerfield *et al.* 2005), including those that rely on agitation of the sample in large volumes of water followed by retention of the community on sieves. Such mass decantation approaches rely on the different settling speeds of abiotic particles compared to the biotic fraction. Medium to coarse grain sediments can often be decanted successfully by mass decantation alone, but muddy or high in clay aggregate samples may require prior rinsing or sonication (Murrell & Fleeger 1989; Giere 2009). Following mass decantation, samples are frequently cleaned using flotation/centrifugation approaches using either sugar solution (Jenkins 1964; Esteves & Silva 1998), or Ludox®, a colloidal silica solution with a specific gravity tailored to user specifications (Markmann & Tautz 2005; Giere 2009). Although passive methods recover both living and dead components of the community, active methods (e.g. Bearmann Funnel and its modifications) differentially recover the most actively moving living components because they depend on organismal locomotion (Baermann 1917; Whitehead & Hemming 1965). During all of these procedures, it is important to note that communities are continually manipulated via the use of measuring cylinders, funnels and stainless steel sieves. From a DNA-based perspective therefore, the potential for cross-contamination of a minor fraction of biodiversity between samples is a concern. Cleaning of apparatus should be rigorous and standardized, with pressurized water augmented by autoclaving and UV treated where possible. Cross-contamination concerns can be tested by performing intermittent negative control experiments, involving no samples.

Experimental design

Case studies

1 Marine littoral benthos: Three 44 mm × 100 mm sediment cores were taken ~10 m apart from the low water intertidal zone from the beach at Littlehampton on the south coast of England, UK, during July, 2007. Samples were stored for *c.* 6 months at room temperature in DESS solution (ratio of 1:3, volume drained sediment to DESS respectively). A meiofaunal fraction designed to include the larger nematodes (45–1000 μm) was isolated by mass decantation, followed by Ludox® (specific gravity 1.16) centrifugation, utilizing combinations of stainless steel sieves and Millipore disposable nylon net filters (Millipore Corporation). Between samples, stainless steel sieves, glass and plasticware were sterilized via combinations of autoclaving (where possible), UV treatment and immersion in 10% sodium metabisulphate solution. Nylon net filters were used to further reduce the risk of cross-contamination, decrease the sample processing time and also remove the biotic community from aqueous solution in anticipation of DNA extraction in lysis buffer.

© 2010 Blackwell Publishing Ltd

2 Tropical rainforest: In March 2007, soil, litter and understory habitats were sampled at La Selva Biological Station, Costa Rica, following the protocol described by Powers *et al.* (2009). Briefly, four locations (at 200, 300, 400 and 500 m markers) along the Sendero Suroeste trail were selected. Within each location, a sampling plot (22 m radius circle, 1520 m^2) was divided into four quadrants. Within each quadrant, one random canopy tree and one random understory tree were selected as sampling points and one soil (15 cm depth) and one litter (overlying soil) samples were collected from 15 cm × 15 cm areas (within 1–2 m away from the canopy and understory trees). A total of eight subsamples (2 trees × 4 quadrants) were pooled to make up one composite soil sample and one composite litter sample per plot. The epiphytic material (e.g. lichen, moss, algae) present on the surface of stems of canopy and understory trees was collected to represent canopy sample. Each tree was sampled at three equidistant (between 2.5 cm and 2.5 m from the soil surface) vertical strata. A 15 × 15-cm area was sampled in each of the strata for a total of 24 subsamples (3 strata × 2 trees × 4 quadrants) pooled to form one composite canopy sample per plot. Samples were stored in a cooler and transported to Universidad Nacional for immediate processing. Because the three habitats were considerably different from the standpoint of the sample matrix (e.g. buoyant organic material of litter vs. non-buoyant particles of soil) and component nematode species (e.g. non-mobile root parasitic nematodes in soil) different extraction methodologies were used to maximize recovery of nematode species from each habitat. Consequently, comparison of habitats *per se* in the example presented here is not appropriate. Litter and canopy samples were cut into smaller pieces, mixed thoroughly and a total of 15–30 g of subsamples were used for nematode extraction. These litter and canopy subsamples were further chopped in a blender in 150 mL of deionized water for 10 s and set onto cotton wool filters (s'Jacob & van Bezooijen 1984) placed in extraction trays. Nematodes were collected at 24- and 48-h intervals and immediately counted for total abundance under an inverted microscope. Nematodes from soils were extracted from ~100 g (standard amount for meaningful nematode analyses) of subsamples using sugar flotation and centrifugation (Jenkins 1964) and counted immediately. To illustrate potential information that can be available from metagenetic studies, data from a single location within the transect (at 200 m: one soil, one litter and one canopy) will be presented.

Community DNA extraction

Following sample manipulation for traditional meiofaunal ecology studies, the community is retained on 25–45 μm stainless steel sieves or filters. Samples are then rinsed from sieves using approximately either 40–60 mL of water, ethanol or DESS, depending on experimental design. However, for the purposes of DNA extraction, all target organisms have to be removed from solution and placed in a suitable cell lysis buffer for DNA extraction. Specimen retrieval can be achieved by removal from Ludox–water interfaces (Markmann & Tautz 2005), centrifugation (but note, the specific gravity of DESS is unsuitable for centrifugation separation), successive subtraction and examination of aliquots of water, or using disposable sieves and 45 μm meshes. The aim is to reduce the community into a volume from which genomic DNA can be effectively liberated.

Once effectively removed from the sample, DNA can be extracted from taxa, but before proceeding, two issues should be considered. First, although extraction methods target organisms of a desired size range, the sample is likely to contain additional taxa such as bacteria, Archaea, Fungi, Plantae, etc., present in the environment, adsorbed on the surface of and present in the guts of targeted organisms. Second, decaying organic matter, containing humic substances and secondary metabolites (e.g. polyphenols, tannins and polysaccharides (Zhou *et al.* 1996; Porebski *et al.* 1997), can potentially inhibit PCR and sequencing reactions.

To achieve effective DNA extraction and overcome the problem of environmentally derived inhibitors, several approaches have been developed to obtain PCR-ready genomic templates from environmental samples. Sample cell disruption can be more effective using bead beating, although there is a risk of shearing DNA into smaller fragments (Picard *et al.* 1992). Conversely, using longer, more gentle treatments, such as spinning wheels, sodium dodecyl sulphate (Huber *et al.* 2002; Sogin *et al.* 2006), enzymes, heat or freeze thaw processes, generally yield higher molecular weight genomic DNA extracts (Zhou *et al.* 1996; Porteus *et al.* 1997; von Wintzingerode *et al.* 1997). Environmental DNA extraction protocols either use a combination of CTAB (cetyltrimethyl ammonium bromide), phenol, chloroform, caesium chloride, etc. (Sambrook *et al.* 1989; Porteus *et al.* 1997) or proprietary chemicals to clean DNA extracts (e.g. Epicentre SoilMaster™, ZR Soil Microbe™ and Mobio PowerSoil™ DNA extraction kits) in association with various column formats. However, given that most proprietary environmental kits are designed for extracting DNA from microbes, they usually have a maximum capacity of ~250 mg of DNA, or eluates of ~200 μL, and it is widely acknowledged that overloading results

© 2010 Blackwell Publishing Ltd

in poor DNA yields. Consequently, for eukaryotic environmental work, either the community has to be partitioned into 250 mg of DNA aliquots or somehow digested in a very low volume lysis buffer prior to kit usage. An alternative may be to use a combination of traditional lysis, followed by a large capacity DNA extraction kit, as used in the marine case study here.

Case studies

1 Marine littoral benthos: After immobilization of the meiofaunal community on disposable nylon net filters, DNA was extracted using the QIAMP DNA Blood Maxi Kit (QIAGEN) following an overnight, spinning-wheel 3 mL proteinase K digestion (Sambrook *et al.* 1989).
2 Tropical rainforest: Samples were transferred into ZR BashingBead Lysis Tubes (Zymo Research Corp) and disrupted using a Mini-BeadBeater (BioSpec Products, Inc.) at maximum speed for 2 min. Genomic DNA was extracted using a ZR Soil Microbe DNA kit according to the manufacturer's protocol.

Choice of genomic loci for delineation of meiofaunal MOTUs

There are clear conceptual differences between metagenetic and specimen-based barcoding analyses. Another difference becomes apparent when choosing genetic marker loci. Hebert *et al.* (2003b) chose the mitochondrial COI gene as the standardized barcoding gene for animals for a number of well-established reasons. Mitochondrial DNA has a haploid mode of inheritance, elevated rate of molecular evolution, lacks introns and has limited recombination (Clayton 1984; Wilson *et al.* 1985; Avise 1994; Piganeau *et al.* 2004; Tsaousis *et al.* 2005). Moreover, indels are rare in mtDNA protein coding genes, universal primers for the COI gene are fairly robust (Folmer *et al.* 1994; Zhang & Hewitt 1997) and the mode of molecular evolution of COI usually facilitates species discrimination while also retaining phylogenetic information for the majority of animal taxa (Hebert *et al.* 2003b). Unfortunately though, the COI gene is not optimal for molecular taxonomic identification purposes for nematodes, because nematode mitochondria have high mutational rates, display excessive saturation, biased substitution patterns and are very A + T rich (Blouin *et al.* 1998; Blouin 2000). Furthermore, primers used for most barcoding studies (Folmer *et al.* 1994) are poorly conserved across nematode diversity, and alternate conserved regions for primer design are not evident (Blouin *et al.* 1998). Parsing and aligning (Chenna *et al.* 2003; Abascal *et al.* 2007) the COI gene from the existing 19 mitochondrial genomes that are currently available for nematodes from NCBI highlights a number of base pair mismatches throughout the LCO1490 and HCO2198 primer annealing regions in the Nematoda, and also a three base-pair deletion in the priming site of LCO1490 in over half of the sequenced taxa. The latter probably explains why previous attempts to use COI in a nematode barcoding framework has resulted in inconsistent amplifications (Bhadury *et al.* 2006). Therefore, acknowledging the lack of universal COI priming sites within nematodes, and the dominance of the Nematoda within meiofaunal communities (Lambshead 2004), suggests that nuclear markers may be more appropriate for meiofaunal metagenetic studies (Blaxter 2003; Blaxter *et al.* 2003).

It is widely acknowledged that alternative markers are required for certain taxa, and attempts are being made to include suites of markers in DNA barcoding. Examples of alternative markers include nuclear ribosomal RNA genes that have been used for decades to identify phyla of microscopic eukaryotes. It was first demonstrated in the 1960s that ribosomal RNA genes (rDNA) and their gene products (rRNA) could be used for the taxonomic classification of microbial species (Doi & Igarashi 1965; Dubnau *et al.* 1965; Pace & Campbell 1971a,b). The genes coding for rRNA are particularly well suited for molecular taxonomy, because they are universally found in all cellular organisms and are of relatively large size. They also contain both highly conserved and variable regions that facilitate the design of very conserved primers that amplify diagnostic regions (Woese 1987; Floyd *et al.* 2002; Markmann & Tautz 2005; Carvalho *et al.* 2009).

For eukaryotes, both the nuclear 18S small subunit (nSSU) and 28S large subunit (nLSU) rDNA genomic regions are excellent candidate genes for molecular identification as they are present in tandemly repeated, multiple copies (50–150 copies/cell), and undergo concerted evolution (Markmann & Tautz 2005). The latter two attributes facilitate their amplification from microscopic organisms that are highly conserved within a species, and divergent among species. DNA barcoding studies utilizing rDNA have focused on the more variable portions of the genes: the D2-D3 'diversity loop' regions of the 28S and the 5' region of 18S. Although both 18S and 28S are probably equally useful for molecular identification, both cases presented here utilize 18S, predominantly because for free-living nematodes at least, universal 18S primer sets are more consistent for PCR amplification than 28S primer sets (Bhadury *et al.* 2006; Porazinska *et al.* 2009). There are also notably more 18S than 28S sequences in public repositories (e.g. recent SILVA databases contain 868 390 18S vs. 143 653 28S entries) (Pruesse *et al.* 2007), enabling more accurate and comprehensive taxonomic assignment to query sequences (Blaxter 2003). What remains less clear is the

© 2010 Blackwell Publishing Ltd

extent to which variation in 18S or 28S genes follows the division of individual organisms into biological species. Within Nematoda, some good species have identical 18S sequences, whereas other congeneric species differ by over 2% (Blaxter *et al.* 1998). 18S (and 28S) are good markers for deep phylogeny (Blaxter *et al.* 1998) but may be less suited to distinguish between closely related taxa. Available data suggest that the D2–D3 loop of 28S may be the better marker in this respect (Ye *et al.* 2007; Subbotin *et al.* 2008).

Marine littoral benthos and tropical rainforest case studies

To select optimal 18S rDNA primer pair combinations, the genomic location of available primers (http://nematol.unh.edu/, http://www.nematodes.org/) and numbers of segregating sites spanning primer pairs ~400 bases apart (recommended for Roche 454 GSFLX sequencing) were investigated using combinations of DnaSP (Librado & Rozas 2009) and MEGA (Tamura *et al.* 2007). Furthermore, the resolving power of target regions (Porazinska *et al.* 2009a) and level of primer sequence conservation across meiofaunal metazoans was also considered (Fig. 1). Consequently, two candidate regions, defined by primers SSU_F04 and SSU_R22 towards the 5′ end (Blaxter *et al.* 1998) and NF1 and 18Sr2b towards the 3′ end of the 18S rDNA (Porazinska *et al.* 2009a) were used independently in the marine and rainforest samples respectively.

PCR and sequencing strategies

At the time of writing, three ultrasequencing platforms (the Roche 454 GS Titanium Series, the Illumina SOLEXA Genome Analyzer and the Applied Biosystems SOLiD™ System) were readily accessible by the research community. Presently, the Roche 454 system is the intuitive choice for any form of metagenetic, or metagenomic analysis, simply because of greater read lengths and subsequent clarity of annotation of individual reads (Blow 2008; Hugenholtz & Tyson 2008). Whereas the Illumina Genome Analyzer and ABI SOLID™ generate many gigabases of sequence data partitioned into 35–75 base reads per instrument run, the Roche 454 Titanium platform generates ~400 Mb of data from 0.8 million 400–450 base reads.

Consensus sequence accuracies of the Roche 454 sequencers range from 99.97% to 99.9984%, with individual per-base error rates of between 0.6% and 0.49% respectively. The large majority of per-base errors (between 39% and 98%) are derived from misreading of the lengths of nucleotide homopolymers effects, including extensions (insertions), incomplete extensions (deletions) and carry forward errors (insertions and substitutions) (Margulies *et al.* 2005; Huse *et al.* 2007). Nucleotide homopolymers are relatively rare in coding genes and in 18S and 28S in particular (M. Blaxter, unpublished).

Following the publication of the Margulies *et al.* (2005) *de novo* sequencing of *Mycoplasma genitalium*, it was clear that parallel pyrosequencing represented a paradigm shift in the cost and volume of sequencing compared to chain-termination (Sanger *et al.* 1977) approaches. However, in order to utilize such sequencing power for multi-sample metagenetic investigations, methods had to be devised in order to pool and then recover amplicons on single, or multiple Roche 454 picotitre plates. Physical gaskets can be used to partition Roche 454 picotitre plates from between 2 and 16 samples (as used in Sogin *et al.* 2006), but this sample multiplicity is inevitably associated with lower overall

Meiofaunal Metazoan		Nematoda
Primer	Sequence	
SSU_F04	5′-G C T T G T C T C A A A G A T T A A G C C-3′	100
% identity	99 96 96 95 98 99 97 97 95 98 99 99 99 99 99 99 99 99 98 98 98	
SSU_R22	5′-G C C T G C T G C C T T C C T T G G A-3′	100
% identity	100 100 99 99 100 100 100 100 100 100 98 99 88 100 97 100 83 99 100	
NF1	5′- G G T G G T G C A T G G C C G T T C T T A G T T-3′	100
% identity	99 100 100 100 100 99 100 100 100 100 100 100 98 100 99 99 100 100 98 100 100 99 99 100	
18Sr2b	5′- T A C A A A G G G C A G G G A C G T A A T-3′	100
% identity	100 99 100 99 100 100 100 100 98 100 100 100 100 100 100 100 92 97 98 95 100	

Fig. 1 The 18S rDNA primer sets used in the marine littoral meiobenthos and tropical rainforest analyses. The number beneath each nucleotide base highlights the conservation of the priming site (calculated by visual inspection) derived from an alignment of ~170 sequences derived from NCBI representing each of the phyla containing meiofaunal representatives (supplied in the Supporting Information). Sequences representing Nematoda were subsampled throughout the currently accepted phylogenetic range presented in Meldal *et al.* (2007) and all base pair positions were 100% conserved in all primer pairs but see Porazinska *et al.* (2009) for further primer bioinformatic comparisons.

© 2010 Blackwell Publishing Ltd

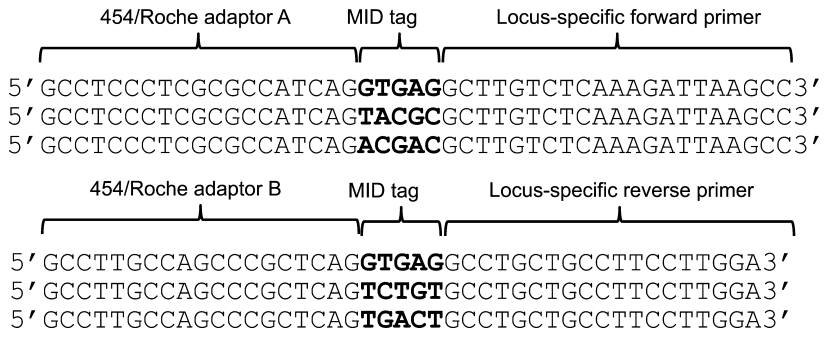

Fig. 2 Examples of three independent primer sets for the identification of mixed metagenetic amplicon pools. The Roche 454 adaptor precedes a five base molecularly identified (MID) tag immediately prior to locus-specific forward and reverse primers respectively. By combining A and B adaptors and appropriate Roche 454 emulsion PCR kits (II or III), bidirectional and unidirectional sequencing can be achieved. In the current example, sequencing from the 5′ end of the forward primer could be performed by hybridizing the Roche 454 adaptor B onto the beads during emulsion PCR and sequencing with the A sequencing adaptor.

sequence throughput per picotitre plate. A number of ways have been suggested to separate samples post-run *in silico*. These range from pooling of easily identifiable different loci (Thomas *et al.* 2006), to use of individually molecularly identified (MID) linkers to independent samples (Meyer *et al.* 2007, 2008; Parameswaran *et al.* 2007; Roche 454). For metagenetic samples, incorporating MID linkers and universal Roche 454 adaptors into fusion primer sets (Binladen *et al.* 2007) is probably the easiest and most cost-effective way of tagging amplicons. Following Binladen *et al.* (2007), study-specific forward primers can be synthesized preceded by a sample-specific MID tag and either Roche 454's A or B universal adaptor sequences (Fig. 2). Thus, each experimental sequence will begin with the MID tag and the PCR primer, and these can be recognized via pattern-matching algorithms to sort individual reads into sample sets. Binladen *et al.* (2007) initially proposed the use of 2 base tags yielding 16 (4^2) different MID combinations. However, Huse (2007) strongly recommends the use of MID adaptors that differ by at least two bases to limit the potential of misallocation due to errors in the MID sequence itself. Eighty-two of the possible 1024 (4^5) five base MID tags that can be combined to fulfil these criteria are currently listed at the Josephine Bay Paul Center's Visualization and Analysis of Microbial Population Structures (VAMPS) website http://vamps.mbl.edu/resources/keys.php, and are supplied in the Supporting Information. For even higher stringency, Hamady *et al.* (2008) and Hamady & Knight (2009) constructed 1544 optimal eight base error-correcting barcodes based on Hamming codes, which minimize redundancy. Given the nature of Roche 454 sequencing, homopolymers are also best avoided and it is optimal (although probably not essential with contemporary read lengths) to utilize combinations of MID adaptors that require the least number of parallel sequencing nucleotide flows (Huse *et al.* 2007; Meyer *et al.* 2007, 2008) to maximize sequencing efficiency through the adapters and primers.

Case studies **1** Marine littoral benthos: The 18S rDNA fragment spanning the primers SSUF04 (5′-GCTTGTAAAGATTAAGCC-3′) and SSUR22 (5′-GCCTGCTGCCTTCCTTGGA-3′) (Blaxter *et al.* 1998) was amplified using MID-tagged fusion primers using 1 μL of genomic DNA template (1:500 dilutions) in a 40-μL reaction using *Pfu* DNA polymerase (Promega), according to manufacturers' recommendations. Sample-specific PCR reactions involved a 2-min denaturation at 95 °C, then 35 cycles of 1 min at 95 °C, 45 s at 57 °C, 3 min at 72 °C and final extension of 10 min at 72 °C. Negative controls were included for all amplification reactions. Electrophoresis of PCR products was carried out on a 2% Top Vision™ LM GQ Agarose (Fermentas) gel and the expected 450-bp fragment was purified using the QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's protocol. All purified PCR products were then quantified with an Agilent Bioanalyser 2100, diluted to the same concentration, pooled and sequenced (A-Amplicon, alongside 10 additional unrelated experimental samples) on a half-plate of a Roche 454 GSFLX sequencer at Liverpool University's Advanced Molecular Genetics Facility, UK.

2 Tropical rainforest: Individual PCR amplifications were performed following protocols described in detail elsewhere (Porazinska *et al.* 2009a) using tagged fusion primers and 1 μL of DNA template. The three metagenetic samples (1 soil, 1 litter and 1 canopy) were sequenced (along with 8 other metagenetic samples from this study) on a single GSFLX half-plate at the Interdisciplinary Center for Biotechnology Research at the University of Florida, Gainesville, FL. Through earlier experiments with artificially assembled nematode communities, it was determined that the use of a single PCR reaction per sample as well as a single emulsion

© 2010 Blackwell Publishing Ltd

PCR (emPCR) and pyrosequencing run was sufficient for both qualitative and quantitative nematode community analysis (Porazinska *et al.* 2009b).

Bioinformatic analyses of metagenetic data sets

The increase in read number and read length generated through contemporary ultrasequencing platforms require novel sequence analysis packages that reduce computational runtime and increase OTU clustering efficiency. Read scaling has made algorithms using direct pairwise comparisons of all available sequences $[N \times (N - 1)]$ computationally intractable (Yu *et al.* 2006; Huson *et al.* 2007) and has reduced the efficiency of programmes using distance matrices methods $[N \times (N - 1)/2]$ (Schloss & Handelsman 2005). Distance matrix methods have been used previously for bacterial metagenetic analyses (Sogin *et al.* 2006). However, the time required for the generation of distance matrices can increase exponentially with an increase in sequence number or metagenetic diversity, and the derivation of a distance matrix inevitably includes estimation of pairwise alignments. Rapid processing of large read numbers requires either reduction of this search space by heuristic avoidance of irrelevant comparisons or implementation of approaches less bound by problems of pairwise comparison. The latter refers to k-mer algorithms, which can cluster sequences based upon the probability of matching a particular word between sequences (Sun *et al.* 2009).

Sequence entry order is a primary concern for developing OTU clustering algorithms. Available programmes assign sequences to OTUs based upon fixed distances from an initial seed (Blaxter *et al.* 2005; Sun *et al.* 2009). Consequently, generation of OTUs can be heavily influenced by an outlier seed. Although it is possible to randomize the sequence entry order, this process becomes increasingly inefficient as progressively more reads are used. Furthermore, randomization of sequence order can lead to variations in the final OTUs (Floyd *et al.* 2002; Blaxter *et al.* 2005).

The analyses that have been performed here are based upon **O**perational **C**lustering of **T**axonomic **U**nits from **P**arallel **U**ltra**S**equencing (OCTUPUS; Sung W, Porazinska D, Creer S, Fonseca VG, Giblin-Davis R, Thomas WK, unpublished), a programme that attempts to address both seeding and runtime problems by interlacing sequence alignments and pairwise comparison in order to generate OTUs (beta version available from the Thomas laboratory on request). OCTUPUS takes advantage of k-mer algorithms (Zhang *et al.* 2000) to make pairwise comparisons against consensus sequences, and can be faster than distance matrix methods (Schloss & Handelsman 2005) or k-mer comparisons using unique sequences (Sun *et al.* 2009). The consensus sequences OCTUPUS uses are continually evolving based upon the sequences assigned to the OTU. Once repeated multiple alignments of the OTU sequences result in an unchanging consensus sequence, the OTU is considered a 'fixed OCTU'. Each OCTU potentially represents one taxonomical group based upon the identity cutoff. By using fixed pairwise comparisons against a variable consensus sequence, seeding error can be reduced.

Data analysis. Sequences generated from the Roche 454 GSFLX from both the marine littoral benthic and tropical rain forest habitats were first checked for quality using Lucy (Chou & Holmes 2001) at default parameters. The sequences were then trimmed, binned according to MID tags and clustered at 95%, 96%, 97%, 98% and 99% similarity match using the OCTUPUS pipeline (Sung *et al.* unpublished). Fixed OCTUs were then compared by MEGABLAST (Zhang *et al.* 2000) against the NCBI database and assigned to phyla if the BLAST similarity match was higher than 90%. A major concern with the analysis of PCR-generated homologous gene regions is the formation of *in vitro* recombinant DNA molecules, or chimeras, where molecules from two different origins artificially combine during PCR (Meyerhans *et al.* 1990). One quick and objective way of flagging a putative chimera is to use the 'greedy' nature of the MEGABLAST algorithm and compare the length of matched bases from the top hit in a MEGABLAST search to the length of the query sequence. As long as the database sequence is longer than the query sequence, and a portion of the 3′ end does not match, it is likely that the query is a recombinant. Given that recombinant molecules can form at any position along a DNA sequence (von Wintzingerode *et al.* 1997; Qiu *et al.* 2001), and referring to previous analyses including control data sets (Porazinska *et al.* 2009a,b), here we apply a strict quality filter allowing a four base length difference between a query OCTU sequence and the matched database sequence for further analyses. Genuine OCTUs that have < 98% coverage with reference sequences will be included within this quality filtering and can be investigated on an individual basis. However here, we prefer to exercise particular caution when extrapolating levels of richness, rather than potentially interpreting patterns of richness that reflect recombinant DNA molecules, instead of real biological entities (Reeder & Knight 2009) .

Data overview and interpretation

Community PCR and sequencing yielded a total of 29 756 high-quality sequences over 200 bases from the marine samples (core 1: 9893, core 2: 9908 and core 3:

© 2010 Blackwell Publishing Ltd

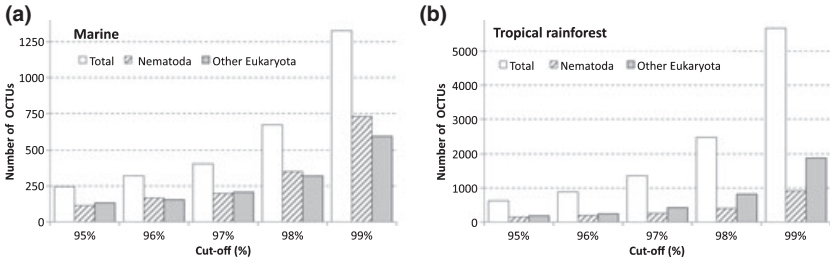

Fig. 3 Number of **o**perational clustering of **t**axonomic **u**nits (OCTUs) found in (a) the marine littoral benthos and (b) tropical rainforest case studies for each base cutoff. Putative non-chimeric OCTU numbers are presented for the total data, Nematoda and other Eukaryota (including OCTUs with BLAST hits to 'environmental samples' representing unclassified taxa).

9955), generating between 246 and 1327 putative non-chimeric OCTUs between the 95% and 99% cutoffs (Fig. 3a). For the tropical rain forest, the three samples yielded a total of 40 334 high-quality sequences of at least 200 bases (soil: 23 742, litter: 10 854 and canopy: 5738), generating between 625 and 5671 putative non-chimeric OCTUs between the 95% and 99% similarity match (Fig. 3b). For convenience, the parsed sequences for the marine and rain forest data sets can be downloaded individually from the NemAToL (http://nematol.unh.edu/) website. Full .sfi files are available from the Short Read Archive at GenBank (SRA010194.2 terrestrial and SRA009394.2 marine). Putative chimera detection for the total data sets ranged from 35% to 38% and 44% to 49% for the F04-R22 (marine) and NF1-18Sr2b (rainforest) data respectively. Many, but not all putative chimeric OCTUs were made up of low copy number reads and accounted for ~20% of reads used in generating the OCTUs.

PCR recombination and chimera formation. The proportion of putative chimeras representing OCTUs is disconcerting, especially as recent control experiments on artificial nematode communities only identified a level of 0.4% of total reads (Porazinska *et al.* 2009a). PCR recombination will suggest the existence of sequences that do not actually exist in the investigated sample, and consequently give a false impression of organismal richness (Markmann & Tautz 2005). *In vitro* recombination of homologous DNA leading to chimeric molecules is widely reported in the microbial literature and detected in databases (von Wintzingerode *et al.* 1997; Qiu *et al.* 2001; Ashelford *et al.* 2005), and levels up to 33% have been reported from meiofaunal communities (Markmann & Tautz 2005). In the latter example, as in many chimera detection approaches (Huber *et al.* 2004; Shaw *et al.* 2008), query sequences were split into 50:50 blast query fragments and were identified as putative chimeras if the 5′ query and the 3′ query had best BLAST matches to different taxa, suggesting that the 5′ and 3′ ends of the sequence are derived from different species. Upon revisiting the data, it is apparent that the 50:50 blast approach can fail to detect some chimeric formations, especially for recombinants that occur further from the sequence midpoint. Accordingly, a stringent base matching approach at the 3′ end of the query sequence against complete reference sequences appears to be the more conservative way of approaching chimera quality control (Porazinska *et al.* 2009b), at least for the taxa involved in this study.

In many ways, metagenetic ultrasequencing experiments are the ideal 'breeding ground' for recombinant DNA molecules. They are based on the amplification of homologous regions from a large number of potentially highly related organisms (von Wintzingerode *et al.* 1997; Qiu *et al.* 2001). The actual species richness and taxonomic composition of the samples may contribute to the level of chimera formation (Qiu *et al.* 2001), but more empirical work needs to be done to assess the level of chimera formation in community-based PCR. The chimera detection approach applied here is particularly aggressive and could also exclude taxa that incorporate five base or more indels in BLAST assignment. Further solutions may therefore be necessary to advance the field of chimera detection, but given that chimeric molecule formation is potentially highly spatially stochastic, a quick and ideal solution may be unattainable without reference to control data sets. It is therefore better to try and reduce the level of DNA recombination within environmental PCRs by adhering to the following procedures (i) performing 'gentle' methods of DNA extraction (enzymatic digestion and using spinning wheels)(Huber *et al.* 2002), rather than bead beating approaches; (ii) increasing polymerase extension times; and (iii) where possible, reducing the number of PCR cycles to the minimum (e.g. 20) (Meyerhans *et al.* 1990; von Wintzingerode *et al.* 1997; Qiu *et al.* 2001).

Sample and taxon coverage. Both sampling and sequencing approaches achieved between 5 and 23 times more sequence coverage per core than is usually revealed with chain termination clone library approaches based

© 2010 Blackwell Publishing Ltd

on ~1000 clone sequences (Kemp & Aller 2004; Blaxter *et al.* 2005; Markmann & Tautz 2005) at ~1% of the cost. The coverage and cost-effectiveness of Roche 454 sequencing therefore brings substantial advantages to studies aiming to elucidate the molecular genetic richness of complex eukaryotic communities.

According to BLAST results, both primer sets (SSU_F04–SSU_R22 and NF1-18Sr2b) amplified homologous 18S gene regions from a substantial proportion of not only meiofaunal, but representatives of the Kingdoms Protista, Plantea and Fungi, in addition to those OCTUs with BLAST hits to 'environmental samples'. The latter are generally representative of the total data (e.g. comprising ~50% nematodes, in addition to further eukaryota) and further manual BLASTing can refine the taxonomic assignment of specific groupings if required. It is clear therefore that both primer sets are very highly conserved in eukaryotes. The primary difference between the two primer sets is that SSU_F04 and SSU_R22 span a more variable region of the 18S rDNA (~30% more polymorphic sites) compared to the NF1-18Sr2b region. Of the meiofaunal phyla that may have been expected to have at least some life history stage present in both environments, notable exceptions are Cnidaria, Nemertea, Rotifera, Brachiopoda and Echinodermata in the marine habitat, and Platyhelminthes, Annelida and Mollusca in the tropical rain forest habitats. Visualizing the conservation of the marine primer sets within the small subunit reference database from SILVA using ARB (Ludwig *et al.* 2004; Pruesse *et al.* 2007) suggests that all of the above, with the exception of cnidarians, should have amplified if genomic DNA was available in the PCR reaction. Therefore, these phyla were either not present in these samples (reflecting actual biology, or the result of taxon extraction methods) or competitive PCR interactions (von Wintzingerode *et al.* 1997) may have prevented amplification of the missing phyla. Revisiting the priming sites of SSU_F04 and SSU_R22, however, reveals that ~50% of cnidarians have a base pair mismatch at the penultimate 3' position of SSU_R22, suggesting that primer mismatching will reduce the amplication of cnidarians in similar studies. In the tropical rainforest case, although undetected in the samples presented here, both flatworms and annelids were recovered in the remaining replicate samples. Absence of molluscs might have been associated with the exclusive nature of the extraction methods.

It is clear from Fig. 3 that OCTU generation at multiple different cutoffs provides very different estimates of richness per sample as OCTUs are created at ever deeper levels of phylogenetic resolution. At fine levels, intraspecific variation will be sampled in some taxa, whereas at deeper levels, certain taxa will be grouping on the basis of genera, order and higher taxonomic levels (Shaw *et al.* 2008). Many nematode morphospecies can be separated on the basis of very low 18S sequence divergence (e.g. 2%) (Blaxter *et al.* 1998), whereas intragenomic and intraspecific sequence variation will invariably be higher in other species and phyla. It is likely that the OCTU cutoff level that broadly correlates with species will occur between 95% and 99% 18S sequence similarity, but there will obviously be exceptions according to the actual species involved in the samples. Without explicitly referring to species, OCTU discrimination does, however, provide comparative metrics that can appraise relative diversity between samples.

Considering the 97% OCTU cutoff, the marine samples two (174 OCTUs) and three (160 OCTUs) contained more than twice the OCTU richness of sample 1 (71 OCTUs) (Fig. 4). Thus, although all three intertidal cores were collected within 10 m of each other minor changes in microhabitat (e.g. sediment grain size, detritus and organic matter and bacterial assemblages) can significantly alter meiofaunal richness between samples, even at microspatial scales (Giere 2009). In the terrestrial data set, although the soil habitat had fewer nema-

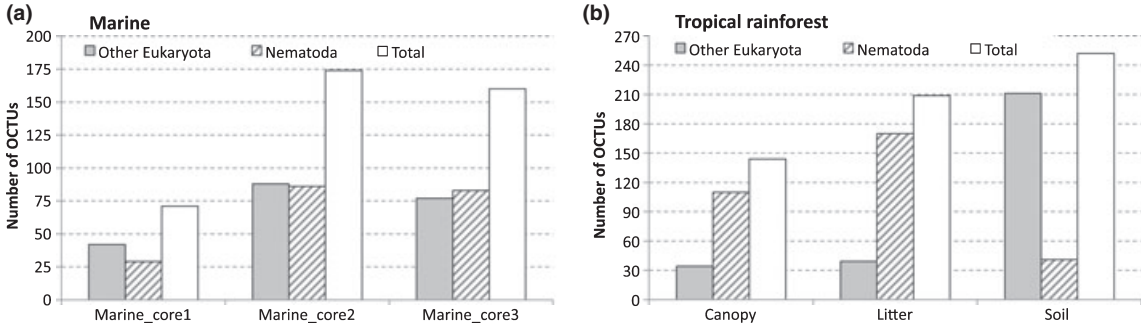

Fig. 4 Number of putative non-chimeric OCTUs (clustered at 97% similarity) found (a) in the marine littoral benthos and (b) tropical rainforest case studies for sample site. Data are provided for totals, Nematoda and other Eukaryota (including OCTUs with BLAST hits to 'environmental samples' representing unclassified taxa).

© 2010 Blackwell Publishing Ltd

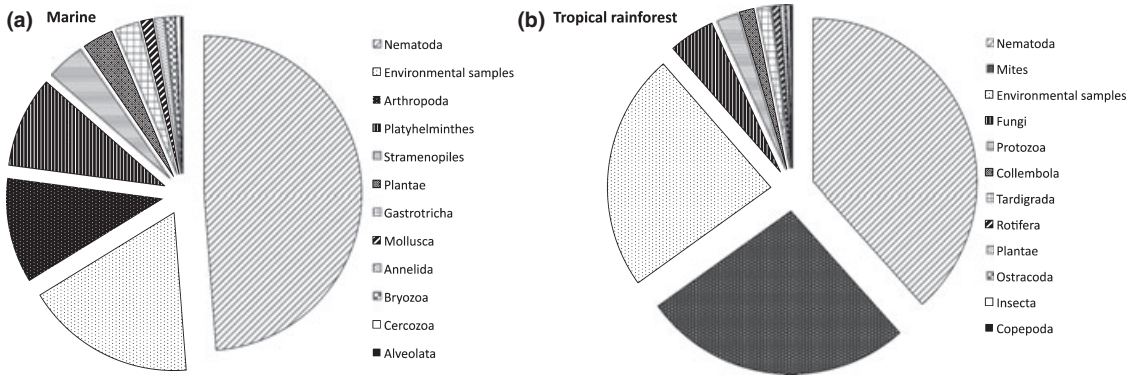

Fig. 5 Pie chart illustrating the relative proportion of OCTUs (clustered at 97% similarity) belonging to each taxonomic grouping found in (a) the marine littoral benthos and (b) the tropical rainforest case studies. BLAST hits to 'environmental samples' represent unclassified taxa.

tode OCTUs (35) than either the litter (149) or canopy (97), the pattern was reversed for other eukaryotes, particularly for mites (soil: 179, litter: 6, canopy: 1) (Figs 4 and 5). As expected, plant-parasitic nematodes were more diverse and abundant in the soil environment, with bacterial- and fungal-feeding nematodes predominating in the litter and canopy. No omnivorous/predatory nematodes were observed in the soil. Recalling that the extraction methods used in the tropical rain forest study were optimized for nematode taxa and also differed between the soil and litter/canopy, the diversity patterns regarding eukaryotes other than nematodes may be inaccurate. For instance, the nearly complete absence of mites and springtails in litter and canopy seems unrealistic. Also, extremely low recovery of fungal sequencing reads is unusual, but there are significant mismatches within the NF1-18Sr2b priming sites for the Ascomycota, Basidiomycota and Glomeromycota primer combinations. Thus, in addition to extraction methods, biology and/or competitive PCR interactions (von Wintzingerode *et al.* 1997), a lack of primer conservation in ecologically relevant groups may explain the lack of fungal sequences in the rainforest data set.

Perspective and future directions in eukaryotic environmental metagenetics

Ultrasequencing accompanied by BLAST annotation can clearly assist in the assessment of relative MOTU richness from large numbers of ecological samples. The increased throughput in sequencing afforded by new-generation sequencers enables faster access (weeks rather than years) to larger amounts of data spanning the breadth of the eukaryotic domain of life, at a fraction of the cost. There are, however, some fundamental limitations regarding the approach. Environmental metagenetics may initially widen the bottleneck and increase the digital sequence flow of information that is currently associated with the taxonomy and identification of smaller organisms in biodiversity assessments (Blaxter 2003; Blaxter & Floyd 2003; Tautz *et al.* 2003). However, without relating species to sequences, the approach will remain analogous to the phylotype approach (Kemp & Aller 2004) adopted for microbial organisms. Relating species to MOTUs will always be a contentious issue, but bioinformatic sequence comparisons and additional molecular assisted taxonomy will inevitably bridge the gap as further voucher specimens are linked to sequences (Blaxter 2003; Blaxter *et al.* 2003; Markmann & Tautz 2005).

The data here refer explicitly to the relative number or richness (McIntosh 1967) of MOTUs and not diversity, which takes into consideration both richness and evenness (Good 1953; Hurlbert 1971; Magurran 2004). As prokaryotes are unicellular and accepting the limitations of PCR-based approaches, microbial ecologists make the assumption that numbers of reads reflect phylotype diversity (i.e. one 16S sequence per individual prokaryote) in metagenetic data sets (Kemp & Aller 2004). The same assumption, however, cannot be readily made for multicellular organisms that comprise different numbers of cells that will change in relation to developmental stage (especially using multicopy 18S markers). Very little work has been performed on the quantitative aspect of environmental metagenetics but preliminary investigations suggest that number of reads may not correlate with small scale differential amounts of DNA template (Binladen *et al.* 2007). Although initial investigations of read abundance in nematode control experiments suggested deviation from hypothesized distribution patterns (Porazinska *et al.* 2009a), repeated experiments clearly indicated that read abundance may be used for quantitative assessments of nematode diversity (Porazinska *et al.* 2009b). Nevertheless, more com-

© 2010 Blackwell Publishing Ltd

prehensive investigations will be required to test the quantitative nature of metagenetic data sets at a range of taxonomic levels. Further to this, trials of independent PCRs (standard and emPCR for Roche 454 sequencing) of the same samples will enable the assessment of replicability within data sets, regarding both quantification and chimera formation.

In addition to overviewing the potential of ultrasequencing in environmental metagenetics, the present data identify clear areas for hypothesis testing and future research directions. Without the constraints associated with chain-termination sequencing and clone library approaches, ultrasequencing will facilitate the investigation of the effects of preprocessing, PCR amplification and sample richness/phylogenetic diversity on chimera formation in a range of genomic loci and experimental scenarios (Meyerhans *et al.* 1990; von Wintzingerode *et al.* 1997; Qiu *et al.* 2001; Lenz & Becker 2008). We also identify two candidate loci within the nSSU that appear optimal for eukaryotic metagenetic analyses, yet questions remain regarding the efficacy of each as an independent marker and work is ongoing to elucidate these issues. It is clear that single primer pair combinations will never co-amplify all taxa and so, in the future, it may be necessary to use primer cocktails (Ivanova *et al.* 2007), or more than one diagnostic region per study, thus providing independent estimates of MOTU richness. Both the nSSU regions here were used due to perceived resolving power and conservation of priming site across the target taxa. However, acknowledging the contemporary ultrasequencing read length available, an explicit comparison with these and other markers (e.g. nLSU and COI; Bhadury *et al.* 2006; Porazinska *et al.* 2009) following primer optimization will quantify which single, or combination, of markers are optimal for eukaryotic metagenetic analyses. The latter will additionally inform which markers will be most appropriate for a reverse taxonomy approach (Markmann & Tautz 2005). Although, as long as molecular voucher specimens/imaging records and DNA extracts exist for focal organisms and/or environmental samples, additional loci can be amplified as and when necessary, thus ensuring homology between data sets.

To conclude, perhaps one of the most crucial issues of environmental metagenetics is to maintain a bioinformatic paper trail so that the scientific community will be able to BLAST annotate future queries and link MOTUs to already existing MOTUs. By linking independent data sets, there is the potential to facilitate the integration of all metagenetic data sets opening up the possibility of ecosystem-based approaches to a range of spatially and temporally heterogeneous evolutionary and ecological questions. We are investigating database mechanisms of how to achieve this and look forward to developments in the emerging field of eukaryotic environmental metagenetics.

Acknowledgements

We would like to thank Neil Hall, Andrew Cossins and Margaret Hughes of the Liverpool Advanced Molecular Genetic Facility for 454 sequencing support, in addition to the staff of Edinburgh University's Gene Pool sequencing service and Holly Bik for field support. Further thanks go to Emlyn Roberts, Wendy Grail and technical support staff at Bangor University. We also thank the Organization for Tropical Studies for access to the La Selva Biological Station sampling plots, Alejandro Esquivel from Universidad Nacional and Thomas O. Powers from the University of Nebraska for technical assistance while in Costa Rica, and Regina Shaw and Bill Farmerie for 454 sequencing support at ICBR, University of Florida. This work was supported by a Natural Environment Research Council (NERC) New Investigator Grant (NE/E001505/1) and a Post Genomic and Proteomics Grant (NE/F001266/1) to SC, an FCT Grant to VGF (SFRH/BD/27413/2006) and a USDA/CSREES–TSTAR Grant 2006-04347 (FLA-FTL-04544) and National Science Foundation (NSF) Biotic Surveys and Inventories projects (DEB-0450537; DEB-0640807) to RMGD.

Conflicts of interest

The authors have no conflict of interest to declare and note that the funders of this research had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Abascal F, Posada D, Zardoya R (2007) MtArt: a new model of amino acid replacement for Arthropoda. *Molecular Biology and Evolution*, **24**, 1–5.

Abebe E, Blaxter M (2003) Comparison of biological, molecular, and morphological methods of species identification in a set of cultured Panagrolaimus isolates. *Journal of Nematology*, **35**, 119–128.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, **71**, 7724–7736.

Avise JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.

Baermann G (1917) Eine einfache Methode zur Auffindung von Ancyclostomum (Nematode) Larven in Erdproben. *Nederlands Tijdscrift Voor Geneeskunde*, **57**, 131–137.

Bhadury P, Austen M, Bilton D *et al.* (2005) Combined morphological and molecular analysis of individual nematodes through short-term preservation in formalin. *Molecular Ecology Notes*, **5**, 965–968.

Bhadury P, Austen MC, Bilton DT *et al.* (2006) Development and evaluation of a DNA-barcoding approach for the rapid

© 2010 Blackwell Publishing Ltd

identification of nematodes. *Marine Ecology Progress Series*, **320**, 1–9.

Bhadury P, Austen MC, Bilton DT *et al.* (2008) Evaluation of combined morphological and molecular techniques for marine nematode (*Terschellingia* spp.) identification. *Marine Biology*, **154**, 509–518.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2 e197**, 1–9.

Blaxter M (2003) Counting angels with DNA. *Nature*, **421**, 122–124.

Blaxter M, Floyd R (2003) Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology & Evolution*, **18**, 268–269.

Blaxter ML, De Ley P, Garey JR *et al.* (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature*, **392**, 71–75.

Blaxter M, Floyd R, Abebe E (2003) Molecular barcoding for nematode identification and diversity studies. *Journal of Nematology*, **35**, 326.

Blaxter M, Mann J, Chapman T *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **360**, 1935–1943.

Blouin MS (2000) Neutrality tests on mtDNA: unusual results from nematodes. *Journal of Heredity*, **91**, 156–158.

Blouin MS, Yowell CA, Courtney CH, Dame JB (1998) Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Molecular Biology and Evolution*, **15**, 1719–1727.

Blow N (2008) Exploring unseen communities. *Nature*, **453**, 687–690.

Carvalho G, Creer S, Allen M *et al.* (2009) *Genomics in the Discovery and Monitoring of Marine Biodiversity*. Springer-Dordrecht; in press.

Chenna R, Sugawara H, Koike T *et al.* (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, **31**, 3497–3500.

Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics*, **17**, 1093–1104.

Clayton DA (1984) Transcription of the mammalian mitochondrial genome. *Annual Review of Biochemistry*, **53**, 573–594.

Coomans A (2000) Nematode systematics: past, present and future. *Nematology*, **2**, 3–7.

Danovaro R, Gambi C, Dell'Anno A *et al.* (2008) Exponential decline of deep-sea ecosystem functioning linked to benthic biodiversity loss. *Current Biology*, **18**, 1–8.

De Ley P, De Ley IT, Morris K *et al.* (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 B: Biological Sciences*, **360**, 1945–1958.

Derycke S, Remerie T, Vierstraete A *et al.* (2005) Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*. *Marine Ecology Progress Series*, **300**, 91–103.

Derycke S, Remerie T, Backeljau T (2008) Phylogeography of the *Rhabditis* (*Pellioditis*) marina species complex: evidence for long-distance dispersal, and for range expansions and restricted gene flow in the northeast Atlantic. *Molecular Ecology*, **17**, 3306–3322.

Doi RH, Igarashi RT (1965) Conservation of ribosomal and messenger ribonucleic acid Cistrons in Bacillus species. *Journal of Bacteriology*, **90**, 384–390.

Dubnau D, Smith I, Morell P, Marmur J (1965) Gene conservation in Bacillus species. I. Conserved genetic and nucleic acid base sequence homologies. *Proceedings of the National Academy of Sciences, USA*, **54**, 491–498.

Edwards RA, Rodriguez-Brito B, Wegley L *et al.* (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*, **7**, 57.

Esteves AM, Silva VMAPD (1998) The behavior of sugar flotation technique in meiofauna extraction from different samples. *Tropical Ecology*, **39**, 283–284.

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839–850.

Floyd RM, Rogers AD, Lambshead PJD, Smith CR (2005) Nematode-specific PCR primers for the 18S small subunit rRNA gene. *Molecular Ecology Notes*, **5**, 611–612.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoeck R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fontaneto D, Kaya M, Herniou EA, Barraclough TG (2009) Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. *Molecular Phylogenetics and Evolution*, **53**, 182–189.

Forest F, Grenyer R, Rouget M *et al.* (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, **445**, 757–760.

Giere O (2009) *Meiobenthology: The Microscopic Motile Fauna of Aquatic Sediments*. Springer-Verlag, Heidelberg.

Good IJ (1953) The population frequencies of species and the estimation of populaton parameters. *Biometrika*, **40**, 237–264.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**, 167–172.

Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. *The Journal of Experimental Biology*, **209**, 1518–1525.

Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Research*, **19**, 1141–1152.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods*, **5**, 235–237.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**, S96–S99.

Higgins RP, Thiel H (1988) *Introduction to the Study of Meiofauna*. Smithsonian Institution Press, Washington, D.C.

© 2010 Blackwell Publishing Ltd

Huber JA, Butterfield DA, Baross JA (2002) Temporal changes in archeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Applied and Environmental Microbiology*, **68**, 1585–1594.

Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon; a program to detect chimeric sequences in multiple alignments. *Bioinformatics*, **20**, 2317–2319.

Huber JA, Welch DBM, Morrison HG *et al.* (2007) Microbial population structures in the deep marine biosphere. *Science*, **318**, 97–100.

Hugenholtz P, Tyson GW (2008) Microbiology: metagenomics. *Nature*, **455**, 481–483.

Hugot J-P, Baujard P, Morand S (2001) Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology*, **3**, 199–208.

Hurlbert SH (1971) The non-concept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.

Huse S (2007) Online commentary on Binladen J *et al.* (2007). *PLoS ONE*, **2**, e197.

Huse S, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research*, **17**, 377–386.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.

s'Jacob JJ, van Bezooijen J (1984) *A Manual for Practical Work in Nematology*. Wageningen Agricultural University, Wageningen, The Netherlands, 77 pp.

Jenkins W (1964) A rapid centrifugal-flotation technique for separating nematodes from soil. *Plant Disease Research*, **48**, 692.

Kemp PF, Aller JY (2004) Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiology Ecology*, **47**, 161–177.

Lambshead PJD (2004) Marine nematode biodiversity. In: *Nematology: Advances and Perspectives. Vol 1: Nematode Morphology, Physiology and Ecology* (eds Chen ZX, Chen SY, Dickson DW), pp. 436–467. CABI Publishing, Wallingford, U.K.

Lambshead PJD, Boucher G (2003) Marine nematode deep-sea biodiversity – hyperdiverse or hype? *Journal of Biogeography*, **30**, 475–485.

Lawton JH, Bignell DE, Bolton B *et al.* (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72–76.

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci. *Gene*, **427**, 117–123.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452. doi: 1410.1093/bioinformatics/btp1187.

Ludwig W, Strunk O, Westram R *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Research*, **32**, 1363–1371.

Magurran AE (2004) *Measuring Biological Diversity*. Blackwell Publishing, Oxford.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

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 B: Biological Sciences*, **360**, 1917–1924.

McIntosh RP (1967) An index of diversity and the relation of certain concepts to diversity. *Ecology*, **48**, 392–404.

Medinger R, Nolte V, Vinay Pandey R, Jost S, Ottenwälder XX, Schlötterer C, Boenigk J (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology* **9** (Suppl. 1), 32–40.

Meldal BHM, Debenham NJ, De Ley P *et al.* (2007) An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. *Molecular Phylogenetics and Evolution*, **42**, 622–636.

Meyer M, Stenxel U, Myles S, Prüfer K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research*, **35**, e96.

Meyer M, Briggs AW, Maricic T *et al.* (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research*, **36**, e5.

Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Research*, **18**, 1687–1691.

Moon-van der Staay SY, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607–610.

Moreira D, Lopez-Garcia P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, **10**, 31–38.

Mou X, Sun S, Edwards RA, Hodson RE, Moran MA (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature*, **451**, 708–711.

Murrell M, Fleeger J (1989) Meiofauna abundance on the Gulf of Mexico continental shelf affected by hypoxia. *Continental Shelf Research*, **9**, 1049–1062.

Pace B, Campbell LL (1971a) Homology of ribosomal ribonucleic acid diverse bacterial species with Escherichia coli and Bacillus stearothermophilus. *Journal of Bacteriology*, **107**, 543–547.

Pace B, Campbell LL (1971b) Homology of ribosomal ribonucleic acid of Desulfovibrio species with Desulfovibrio vulgaris. *Journal of Bacteriology*, **106**, 717–719.

Parameswaran P, Jalili R, Tao L *et al.* (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research*, **35**, e130.

Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. *Ecology Letters*, **9**, 741–758.

Picard C, Ponsonnet C, Paget E, Nesme X, Simonet P (1992) Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Applied and Environmental Microbiology*, **58**, 2717–2722.

Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, **21**, 2319–2325.

Platt HM, Warwick RM (1983) *Free-living Marine Nematodes. Part I British Enoplids*. Cambridge University Press, Cambridge.

Porazinska D, Giblin-Davis R, Faller L *et al.* (2009a) Evaluating high-throughput sequencing as a method for metagenomic

© 2010 Blackwell Publishing Ltd

analysis of nematode diversity. *Molecular Ecology Resources*, **9**, 1439–1450.

Porazinska DL, Sung W, Giblin-Davis RM, Thomas WK (2009b) Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. *Molecular Ecology Resources*, doi: 10.1111/j.1755-0998.2009.02819.x.

Porebski S, Bailey L, Baum B (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, **15**, 8–15.

Porteus LA, Seidler RJ, Watrud LS (1997) An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. *Molecular Ecology*, **6**, 787–791.

Powers TO, Neher DA, Mullin P *et al.* (2009) Tropical nematode diversity: vertical stratification of nematode communities in a Costa Rican humid lowland rainforest. *Molecular Ecology*, **18**, 985–996.

Pruesse E, Quast C, Knittel K *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

Qiu X, Wu L, Huang H *et al.* (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, **67**, 880–887.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Reeder J, Knight R (2009) The 'rare biosphere': a reality check. *Nature Methods* **6**, 636–637.

Rodriguez-Brito B, Rohwer F, Edwards RA (2006) An application of statistics to comparative metagenomics. *BMC Bioinformatics*, **7**, 162.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sanger F, Nicklen S, Coulsen AR (1977) DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Science, USA*, **74**, 5463–5467.

Schloss P, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, **71**, 1501–1506.

Seutin G, White B, Boag P (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, **69**, 82–90.

Shaw AK, Halpern AL, Beeson K *et al.* (2008) It's all relative: ranking the diversity of aquatic bacterial communities. *Environmental Microbiology*, **10**, 2200–2210.

Snelgrove P, Blackburn TH, Hutchings PA *et al.* (1997) The importance of marine sediment biodiversity in ecosystem processes. *Ambio*, **26**, 578–583.

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored ''rare biosphere''. *Proceedings of the National Academy of Science, USA*, **103**, 12115–12120.

Somerfield PJ, Warwick RM, Moens T (2005) Meiofauna techniques. In: Eleftheriou A, McIntyre A (eds) *Methods for the study of the marine benthos*. Blackwell, Oxford, pp. 229–272.

Stoeck T, Bass D, Nebel M *et al.* (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* **19** (Suppl. 1), 21–31.

Subbotin SA, Ragsdale EJ, Mullens T *et al.* (2008) A phylogenetic framework for root lesion nematodes of the genus Pratylenchus (Nematoda): evidence from 18S and D2-D3 expansion segments of 28S ribosomal RNA genes and morphological characters. *Molecular Phylogenetics and Evolution*, **48**, 491–505.

Sun Y, Cai Y, Liu L *et al.* (2009) ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Research*, **37**, e76, 10.1093/nar/gkp285.

Sutherland W, Armstrong-Brown S, Armsworth P *et al.* (2006) The identification of 100 ecological questions of high policy relevance in the UK. *Journal of Applied Ecology*, **43**, 617–627.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.

Thomas W, Vida J, Frisse L, Mundo M, Baldwin J (1997) DNA sequences from formalin-fixed nematodes: integrating molecular and morphological approaches to taxonomy. *Journal of Nematology*, **29**, 250–254.

Thomas RK, Nickerson E, Simons JF *et al.* (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Medicine*, **12**, 852–855.

Tsaousis A, Martin D, Ladoukakis E, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, **22**, 925–933.

Venter JC, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.

Warwick RM, Somerfield PJ (2008) All animals are equal, but some animals are more equal than others. *Journal of Experimental Marine Biology and Ecology*, **366**, 184–186.

Whitehead A, Hemming J (1965) A comparison of some quantitative methods of extracting small vermiform nematodes from soil. *Annals of Applied Biology*, **55**, 25–38.

Wilson AC, Cann RL, Carr SM *et al.* (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, **26**, 375–400.

von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213–229.

Woese CR (1987) Bacterial evolution. *Microbiological Reviews*, **51**, 221–271.

Ye W, Giblin-Davis R, Braasch H, Morris K, Thomas W (2007) Phylogenetic relationships among Bursaphelenchus species (Nematoda: Parasitaphelenchidae) inferred from nuclear ribosomal and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **43**, 1185–1197.

Yoder M, Ley ITD, King IW *et al.* (2006) DESS: a versatile solution for preserving morphology and extractable DNA of nematodes. *Nematology*, **8**, 367–376.

© 2010 Blackwell Publishing Ltd

Yu Y, Breitbart M, McNairnie P, Rohwer F (2006) FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics*, **7**, 57.

Zhang D-X, Hewitt GM (1997) The long and short of nuclear mitochondrial DNA (Numt) lineages – reply. *Trends in Ecology and Evolution*, **12**, 114.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, **62**, 316–322.

SC is a molecular ecologist with broad interests in the spheres of ecology and evolution. Current research foci include understanding the mechanisms underpinning the role, origins and dynamics of biodiversity in ecosystem processes, environmental genomics, molecular phylogenetics and population genetics. VGF's research interests focus on functional and environmental genomic aspects of differential gene expression under normal and pathological conditions, the analysis of marine biodiversity from a taxonomic and evolutionary perspective, and ecology and management of marine ecosystems. DP and RGD are interested in the patterns of terrestrial microbiotic diversity and their role in ecosystem functioning using ultrasequencing approaches as well as in molecular phylogeny of terrestrial and entomophilic nematodes for linking metagenetics with traditional taxonomy. DMP is interested in molecular evolution traced using gene families, particularly those regulating whole animal homeostasis. MP is an ecologist specializing in marine benthic biodiversity. PJD is a nematologist interested in the quantitative analysis of biodiversity and its origins. GRC's research is aimed at the elucidation of fundamental aspects of a species' biology such as patterns of dispersal and gene flow, evolution of life histories and behaviour, response to environmental stress, and mechanisms of speciation. MB is a genomics biologist, with major interests in the genomics and molecular identification of ''neglected'' animal phyla. WKT and WS are interested in patterns and processes of genomic change and how these relate to our understanding of genetic diversity and evolution, in addition to the analysis and interpretation of metagenetic data.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data 1. List of potential barcode tag primer keys currently used for environmental metagenetic work from the Josephine Bay Paul Center's Visualization and Analysis of Microbial Population Structures (VAMPS) web resource (http://vamps.mbl.edu/resources/keys.php).

Data 2. Per cent similarity matches for the studied primers were deduced using up to 13 representative sequences from each Metazoan taxon containing meiofaunal representatives from GenBank sequence database for nSSU sequences.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2010 Blackwell Publishing Ltd