Molecular Ecology (2002) **11**, 839–850

Molecular barcodes for soil nematode identification

ROBIN FLOYD, EYUALEM ABEBE, ARTEMIS PAPERT and MARK BLAXTER
Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

Abstract

Using a molecular barcode, derived from single-specimen polymerase chain reaction (PCR) and sequencing of the 5′ segment of the small subunit ribosomal RNA (SSU) gene, we have developed a molecular operational taxonomic unit (MOTU) scheme for soil nematodes. Individual specimens were considered to belong to the same MOTU when the sequenced segment of 450 bases was > 99.5% identical. A Scottish upland *Agrostis-Festuca* grassland soil was sampled, using both culture-based and random selection methods. One hundred and sixty-six cultured isolates were sequenced, and clustered into five MOTU. From 74 randomly sampled individuals across the study site, 19 MOTU were defined. A subsequent sample of 18 individuals from a single subplot contained eight MOTU, four of which were unique to the single subplot sample. Interestingly, seven of these MOTU were not present in the culture-independent sampling. Overall, a total of 23 MOTU were defined from only 240 sequences. Many MOTU could readily be assigned to classical, morphologically defined taxonomic units using a database of SSU sequences from named nematode species. The MOTU technique allows a rapid assessment of nematode taxon diversity in soils. Correlation with a database of sequences from known species offers a route to application of the technique in ecological surveys addressing biological as well as genetic diversity.

Keywords: biodiversity assessment, DNA sequence, nematodes, 18S ribosomal RNA (SSU)

Received 17 August 2001; revision received 7 December 2001; accepted 7 December 2001

Introduction

Measurement of meiofaunal diversity and abundance is an important but time consuming process. Morphological identification of individual organisms to named species is often not technically possible due to sheer abundance, small size, and lack of expert knowledge of the groups encountered. This is especially true of nematodes, whose diversity in soils and sediments remains essentially unknown. Surveys of benthic sediments suggest that the total species number for marine nematodes may exceed 1 million (Lambshead 1993; Lambshead 2001), with only a few thousand described in the scientific literature (Malakhov 1994; De Ley & Blaxter 2001). In terrestrial systems, nematode diversity appears to be under-reported (Lawton *et al*. 1998), with, for example, only about 200 species of soil nematodes being described from the British Isles (Boag & Yeates 1998). The maximum number of nematode taxa described from a single soil site is 228 from a prairie in Kansas, USA (Orr & Dickerson 1966; Boag & Yeates 1998). Given that many (or most) nematode species have yet to be formally described morphologically (Platt 1994), a robust and transferable system of identification, applicable to all individuals and taxa, is sorely needed.

As terrestrial nematodes can easily exceed one million individuals per square metre of soil, it is likely that any attempt to exhaustively describe a local nematode fauna will become an undertaking of monographic proportions. In addition, many taxa can be diagnosed only from adult male- or female-specific structures, or from population measures of relative morphological characters. In such cases, larvae, individuals of the 'wrong' sex, or individual specimens may not be identifiable. For many studies, identifications are only made to generic level, and taxa are designated as 'genus_x 1', 'genus_x 2'. This precludes simple correlation of surveys carried out by different experts at different sites and times.

We approach this problem from a use-value perspective. We would like to develop a method that is simple, universal and cross-compatible between surveys. We aim to define operational taxonomic units (OTU) relevant to the study at hand. These OTU need not have any formal correlation with published species descriptions, though such

Correspondence: Robin Floyd. Fax: +44 131650 7489; E-mail: Robin.Floyd@ed.ac.uk

correlation could be achieved, and their definition should remove the need for explicit identification to species level. However, with meiofaunal organisms such as nematodes (most of which are less than 1 mm in length) the paucity and microscopic size of easily discerned distinguishing morphological characters makes application of an OTU approach using morphology onerous (Lawton *et al*. 1998). In addition, the question of how to achieve between-sample, between-experiment and between-laboratory comparison of OTU remains problematic. Universal acceptance of an agreed character scoring scheme would allow the use of morphology, but might run into problems when taxa with previously unrecorded character states or character combinations are found.

A genetic profile, or molecular barcode, derived from the nuclear or mitochondrial genome of the individuals studied, might overcome these difficulties. Using molecular markers that are stable within experimental time, diagnostic of experimentally relevant OTU, and can be described rigorously, it should be possible to define molecular operational taxonomic units (MOTU). Such molecular barcodes should be applicable to all life cycle stages.

Molecular methods for diversity assessment have already aided understanding of groups of organisms that are difficult or impossible to study by other means. The application of culture-independent methods of taxonomy to bacterial flora has revealed unexpected diversity in most habitats. For example, 70% of PCR-amplified eubacterial 16S genes from Siberian tundra soil differed by 5–15% from those in current databases, and a further 7% differed by more than 20% from known sequences (Zhou *et al*. 1997). It was concluded that the majority of the tundra soil bacterial community had never been isolated, and that the physiology and function of its dominant members was unknown. Analysis of the sequenced, culture-independent bacterial diversity suggests that only 1% of diversity may be culturable (Woese 1996), and that there exist widespread and ecologically important major groups (bacterial divisions) for which no cultured isolates are available (Hugenholtz *et al*. 1998). While it is unlikely that a meiofaunal group such as nematodes has been similarly undersampled, it remains likely that a majority have yet to be described, and it is certain that only a tiny minority have any associated sequence data.

Several molecular fingerprint systems have been proposed and tested for nematodes, including length polymorphism in polymerase chain reaction-amplified gene segments, restriction fragment length polymorphisms (RFLP), randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP) (Powers 1992; Powers & Harris 1993; Powers & Adams 1994; Folkertsma *et al*. 1996; Powers *et al*. 1997; Szalanski *et al*. 1997; Semblat *et al*. 1998; Semblat *et al*. 2000). These approaches have significant drawbacks, however. PCR and RFLP are only applicable to a small subset of known taxa, as the methods display only a limited amount of information (the presence and length of PCR-amplified DNA and restriction enzyme fragments). RAPD and AFLP analyses can display huge amounts of information (hundreds of fragments), but it remains unclear what level of difference in fragment patterns should be taken as defining an OTU. In all of these methods, when a novel pattern is observed there is no simple way of deducing the relationship of the individual from which it derives, to known previously described taxa. Molecular sequence data has been used several times to define taxa of nematodes. Sequences from the nuclear ribosomal RNA repeat have been used to demonstrate the probable identity of isolates from different parasitic hosts (Elson-Riggins *et al*. 2001), and to unravel the relationships of species complexes that suffer from confused published taxonomy (Adams 1998; Adams *et al*. 1998; Beckenbach *et al*. 2000).

We are endeavouring to develop a simplified, molecular system that will permit diversity and abundance estimation of nematodes in soils and elsewhere using a standardized methodology applicable in all situations. We report here on our first steps towards this system, based on soil nematode surveys carried out on the UK Natural Environment Research Council (NERC) Soil Biodiversity and Ecosystem Function study site, Sourhope farm in Southern Scotland. We demonstrate that PCR, sequencing and analysis of an informative DNA segment of the small ribosomal subunit RNA gene is a powerful tool for determining, quantifying and interpreting MOTU of soil nematodes.

Materials and methods

Study site and sampling regime

Our study site was at Sourhope farm, near Kelso, in the Scottish Southern Uplands, abutting the English-Scottish border (grid reference NT 620 384). The site is a hill farm grassland ecosystem (altitude ~260 m) dominated by *Agrostis* and *Festuca* species (soil type U4 in the UK soils classification). The site is the subject of a wide-ranging co-ordinated study of soil biodiversity (for additional details of the site see the Soil Biodiversity and Ecosystem Function in Soil Programme website at http://mwnta.nmw.ac.uk/soilbio/index.html), and is divided into control and experimental perturbation plots. Grazing animals have been excluded from the site since 1998. All samples were taken from five undisturbed control plots in the summers of 1999 and 2000. Soil on the site was sampled to a depth of 10–15 cm. A 2.5 cm diameter soil corer was used. Each core was divided into an upper, organic rich horizon, and a lower mineral horizon of approximately 5 cm each. Soil samples were stored at 4 °C until used.

Nematode isolation

Nematodes were isolated from soil samples by a standard filter extraction procedure (Southey 1986). While this method does not extract all nematodes, it is fast and repeatable. Soil was spread thinly over one layer of Kimberly Clark lab tissue suspended over 0.5 cm of sterile tap water by a wide mesh filter. After 18–24 h at 15 °C, nematodes that had migrated into the water were collected by centrifugation. For morphological identification, nematodes were fixed in hot ~60 °C 4% formaldehyde and transferred to anhydrous glycerine according to the method of Seinhorst (Seinhorst 1959) as modified by De Grisse (De Grisse 1969). Permanent slides were prepared according to Cobb (Cobb 1918). We used Zeiss Axiovert and Olympus BX 50 microscope to study all specimens.

Culturing

Randomly selected individual adult female nematodes were picked onto 20% Modified Youngren's Only Bactopeptone (MYOB) agar plates (per 10 L: 1.1 g Tris-HCl; 0.48 *g* Tris base; 6.2 *g* peptone; 4 *g* NaCl; 16 mg cholesterol; 210 *g* agar), seeded with *Escherechia coli* OP50, and cultured at 15 °C. Plates were monitored weekly for up to six weeks to identify nematodes that founded cultures. No particular effort was made to exclude bacterial and fungal carry-over from the soil. Established cultures were maintained by passage on 20% MYOB/*E. coli* plates. Some cultures were isolated from primary plates supplemented with small pieces of potato tuber. While some strains could be cryopreserved at −80 °C, most did not survive freezing, and were maintained by serial passage. Each monoculture was allocated a unique six-character ID code, following the nematode genetic nomenclature guidelines (Bird & Riddle 1994). All Sourhope cultures have been allocated sequential codes beginning from ED2000.

Choice of DNA marker for MOTU discovery

In considering which segment of DNA to use for generating a molecular barcode, issues of both diversity and conservation are relevant. Diversity of the chosen sequence segment between relevant taxa (for example morphologically recognised species) is necessary in order to be able to define unique sequences corresponding to the diversity. Conservation of sequence (or at least flanking regions of the sequence) is necessary in order to be able to use universal PCR primers. Conservation within the sequence segment aids in alignment of sequences from different MOTU, and thus putative identification of otherwise anonymous specimens by comparison to sequences from named taxa. We examined the ribosomal RNA (rRNA) gene repeat as a possible source of barcode sequence. While the internal transcribed spacer (ITS) regions are highly divergent between taxa, and are flanked by conserved primer sites in the coding rRNAs, it is diffcult to align ITS regions between disparate taxa, and within-species variation in ITS length and sequence has been observed in diverse nematodes. The small subunit rRNA (SSU or 18S) sequence dataset for nematodes is currently unique for the phylum because sequences are available for a large number of identified specimens from across the known phylogenetic diversity (Blaxter *et al*. 1998; Dorris *et al*. 1999). The 5′ third of the ~1600 base pair SSU gene contains about 50% of the nucleotide variability of the whole gene, as it encompasses both conserved stem and highly divergent loop regions. This pattern of conservation and divergence recommended it for analysis, as the gene is of a relatively constant length, and can be aligned with some confidence. The SSU gene is present in 50–100 copies per genome, and thus is a more abundant target than a single copy gene. We thus chose the SSU gene for these initial studies.

Single nematode digestion and PCR

Individual nematodes (adults and larvae) were picked directly into 20 µL of 0.25 M NaOH in 0.2 mL tubes, then kept at room temperature for 3–16 h (Stanton *et al*. 1998). This lysate was then heated for 3 min at 95 °C. 4 µL of HCl and 10 µL of 0.5 M Tris-HCl buffered at pH 8.0 were added to neutralize the base. 5 µL of 2% Triton X-100 was also added, and the lysate was heated for a further 3 min at 95 °C. Lysates were stored at –20 °C.

For PCR, 0.5–2 µL of each lysate was added to a 50-µL PCR reaction in a microtitre plate comprising Expand LT buffer 3 at 1 × concentration; 2.25 mM $MgCl_2$; 0.2 mM each nucleotide; 1.3 units of Expand LT polymerase (Roche Biochemicals); and 75 ng each primer. The primers used were SSU18A (AAAGATTAAGCCATGCATG) and SSU26R (CATTCTTGGCAAATGCTTTCG) (Blaxter *et al*. 1998), giving a ~1000 bp PCR product. The reaction conditions were: 94 °C for 5 min; 35 cycles of {94 °C for 1 minute; 52 °C for 1 minute 30 s; 68 °C for 2 min}; 68 °C for 10 min Products (5 µL) were visualized on agarose gels stained with ethidium bromide.

PCR-available DNA was released in as little as 1 h in 20 mM NaOH, but the optimal time for digestion was between 3 and 16 h. Over-digestion gave poorer results (less strong and/or less frequent bands). In general, 2 µL of NaOH digest could be used in a 50-µL PCR reaction. 1 µL digest per 50 µL PCR also gave product in some cases, but less reliably (1 µL may provide sufficient DNA if the nematode is large, but not if it is small, whereas 2 µL provides enough in all cases). Therefore, a single 39 µL nematode digest provides sufficient DNA for between 20 and 40 PCRs.

DNA Sequencing

Successful PCRs were treated directly with exonuclease I and shrimp alkaline phosphatase to remove primers and nucleotide triphosphates (3 μL SAP and 4.5 μL *Exo*I were added to 45 μL PCR product; reactions were heated at 37 °C for 40 min and 94 °C for 15 min), and 2 μL of the cleaned PCR product taken to an Applied Biosystems BigDye sequencing reaction (10 μL reaction volume) using the primers SSU18A or SSU9R (AGCTGGAATTACCGCGGCTG) (Blaxter *et al*. 1998). Reactions were electrophoresed and sequence chromatograms collected on an Applied Biosystems 377 sequencer.

For sequencing the 5′ 500 base pairs, we initially used SSU18A, the 5′ primer used for PCR. However, in some cases this gave poor quality sequence data. We therefore used the primer SSU9R, which anneals in the reverse orientation 500 base pairs into the molecule, for routine sequencing. SSU9R gave more robust results than SSU18A. From a reasonably strong and clean PCR product, we reliably obtained 450–500 bases of high quality sequence.

Single nematodes picked directly from soil samples (not grown in culture) were given unique numbers, using a system with five digits beginning at 10 000, followed by 'ED', so that these could be easily distinguished from cultured nematodes.

Cluster Analysis

Sequence traces were automatically trimmed of poor quality data using PHRED (Ewing & Green 1998; Ewing *et al*. 1998), and aligned to each other using CLUSTALX (Thompson *et al*. 1997; Jeanmougin *et al*. 1998). For MOTU clustering we aggressively removed from the aligned dataset all ambiguous characters (such as gaps, and unresolved base calls). The elimination of this potentially noisy data was carried out to avoid treating base-calling errors as significant, and also to eliminate regions that had alignment problems (and were thus characterized by frequent insertion of gaps). While this process necessarily removed some phylogenetically informative data, it also avoided the use of questionable characters. The alignments were processed to predict MOTU content using the neighbour joining algorithm, with absolute character differences as a distance measure (i.e. no corrections for transition vs. transversion, and no correction for multiple substitution), in PAUP* 4.0b6 (Swofford 1999; Swofford *et al*. 1996). For analyses investigating the relationships between MOTU and sequences from known taxa, the neighbour-joining algorithm was used with Kimura two-parameter distance and proportion of variable sites corrections.

Accuracy of sequencing

To examine experimental error, we subjected eight cultured nematode isolates to multiple resequencing. Twelve individuals of each isolate were picked, digested, amplified and sequenced using the standardized protocol. Analysis of the resulting sequences showed that the sequencing error was 1 or 2 bases in 500 aligned characters (i.e. in each group of 12 sequences, 10 or 11 were identical, while one or two typically contained a base difference), affirming the cut-off defined above for definition of each cluster (data not shown). We thus designate a MOTU as a cluster of sequences that differ from each other by less than three bases over the aligned and analysed region. We chose not to perform resequencing, or double-stranded sequencing, of the PCR products as we wished to develop a high-throughput and relatively cheap method.

Results

A robust method for single nematode PCR of the 5′ end of the SSU gene

NaOH digestion followed by SSU PCR and sequencing of individual nematodes had an 80–85% success rate. Alternate methods, involving proteinase K digestion or simple lysis, were much less robust (data not shown). We could detect no phylogenetic bias in the sequences generated, as they originate from across the known diversity of nematodes (see below). There was no clear correlation with size or stage of nematode and success rate. Repeated trials yielded PCR products and sequence for all the cultured isolates. The retention of ~80% of the DNA extract from each nematode permits repeated attempts at amplification of the same segment, or amplification of multiple segments from the same specimen. The DNA extract, in buffered solution, can be frozen at −80 °C and kept as a voucher for the specimen.

Sampling of cultures and verification of accuracy of sequencing for MOTU assignment

Twelve hundred individual nematodes were transferred to culture plates, and 166 were established as monocultures. These cultures were each identified to species (or, in some cases, to genus only when 'difficult' genera were present or the particular pattern of morphological characters did not accord with described species), and five morphological taxa were found (Table 1). Individual nematodes from each culture were sequenced and the sequences analysed for MOTU content as described. Five different MOTU can be derived from the 166 sequences sampled (Fig. 1). MOTU and morphological taxon assignments agree

Table 1 Nematode cultures and MOTU from Sourhope

MOTU	Morphological identification	Number of independent cultures
MOTU_culture_1	*Pellioditis* sp.	22
MOTU_culture_2	*Pristionchus lheritieri*	7
MOTU_culture_3	*Acrobeloides* sp.	132
MOTU_culture_4	*Panagrolaimus* sp.	3
MOTU_culture_5	*Panagrolaimus* sp.	2
		166

for all cultures, except for MOTU_culture_4 and MOTU_culture_5 (see discussion). As further confirmation of the robustness of the MOTU system, all the cultures morphologically identified as *Acrobeloides* sp. were within 2 base pairs of each other on the neighbour joining analysis.

Random sampling of untreated plots across site and assessment of nematode diversity by MOTU

Seventy-four high-quality sequences were generated from randomly picked nematodes from across the five control plots on the Sourhope field site. Nineteen clusters of sequences were identified within which sequences differ by less than 3 nucleotides over the included characters (Fig. 2). These have been designated MOTU_sample_1 to MOTU_sample_19. A subsample of 18 sequences from one subplot (subplot 4 DU) yielded 8 MOTU (Fig. 3), 4 of which were unique to the single-site sample. Our current random survey total of 23 MOTU is likely to be a significant underestimate of the real (molecular) diversity of nematodes at Sourhope.

From the two datasets (cultured and random) we generated majority-rule consensus sequences for each

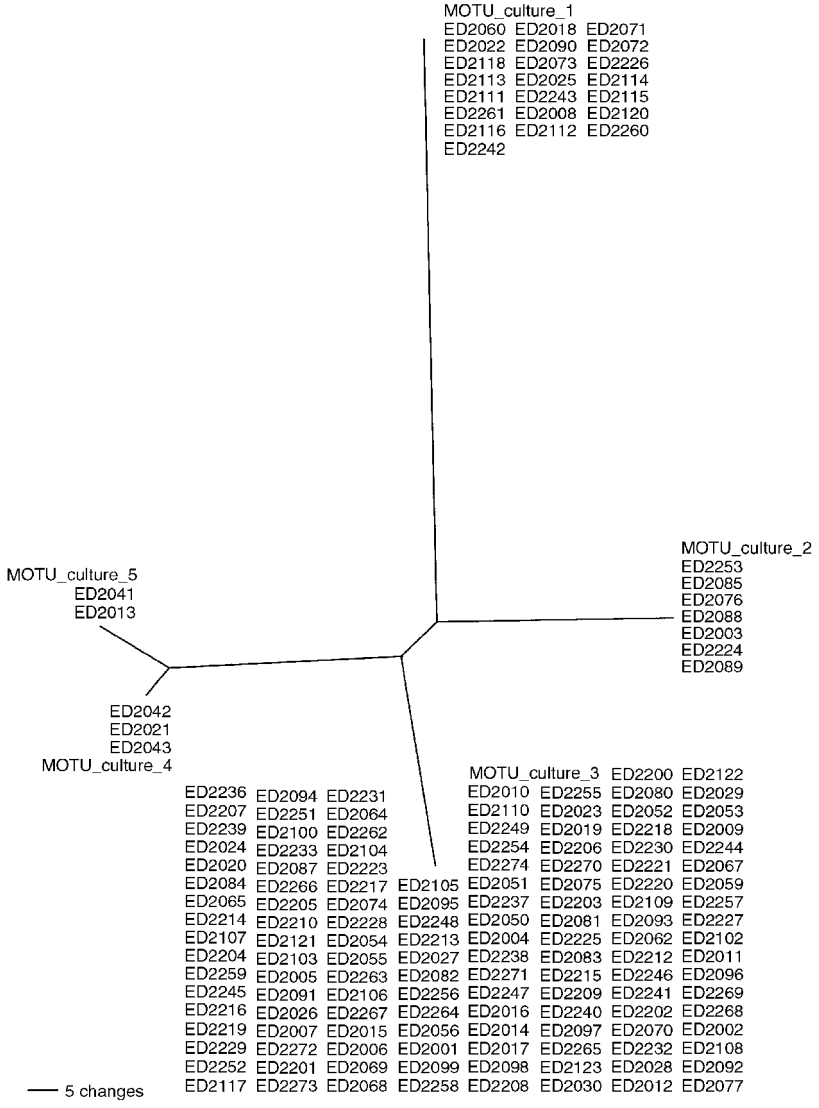

Fig. 1 Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from cultured nematode isolates. One hundred and sixty-six sequences from nematode cultures initiated from single specimens were aligned and analysed as described in materials and methods. The analysis included 349 of the aligned nucleotides. The resultant tree is here represented as an unrooted phylogram, with branch lengths corresponding to those estimated from the uncorrected neighbour joining analysis (with missing and gapped sites excluded). Each cluster of sequences, identified by their specimen code, is designated with a MOTU number.

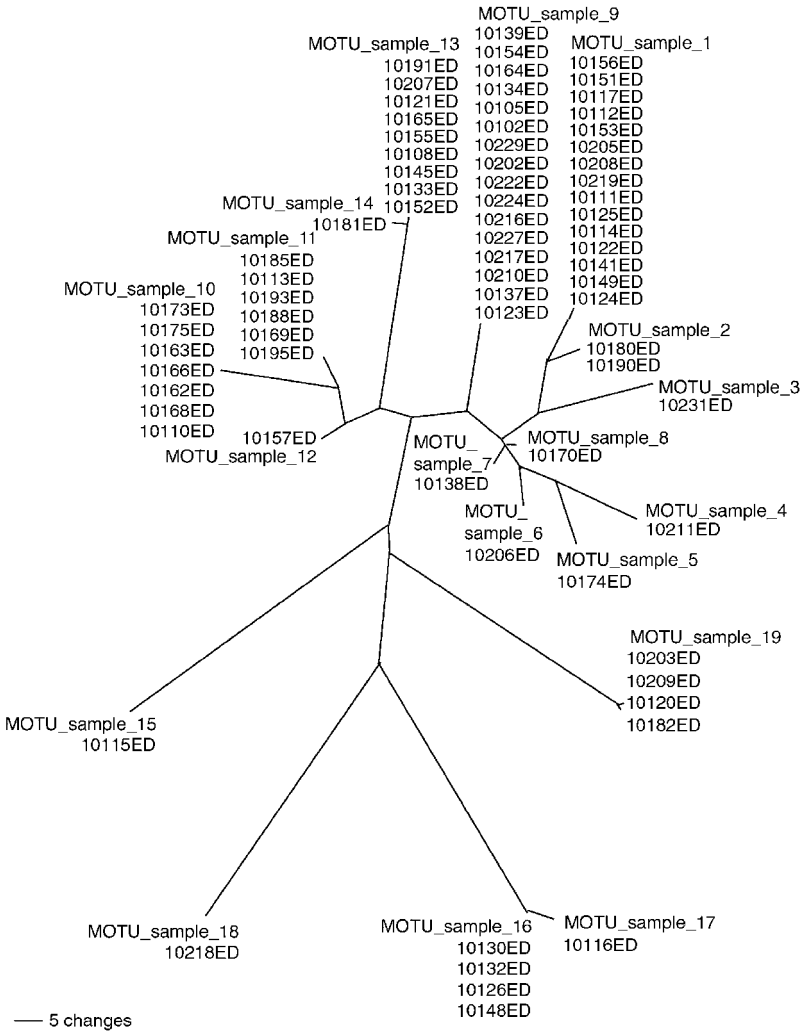

Fig. 2 Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from a random sample. Seventy-four sequences derived from single nematode specimens across the Sourhope field site were aligned and analysed as described in materials and methods. The analysis included 350 of the aligned nucleotides. The resultant tree is here represented as an unrooted phylogram, with branch lengths corresponding to those estimated from the uncorrected neighbour joining analysis (with missing and gapped sites excluded). Each cluster of sequences, identified by their specimen code, is designated with a MOTU number.

MOTU, and aligned them to a selection of sequences from identified nematode species. The named nematode sequences were selected on the basis that they were the closest matches (in sequence similarity analysis) to one or more of the MOTU consensus sequences. The resultant phylogram (Fig. 4) allows us to compare the MOTU found in each sample and sequences from named nematodes.

The 5 MOTU from cultured isolates correspond to one sample MOTU and four MOTU only seen in cultures. The culture sample is derived from a screen of 1200 nematodes and thus we would expect to observe these sequences in an enlarged random screen.

Using sequences from known taxa as comparators we can assign MOTU to described nematode taxa (Fig. 4). For example, very robust assignments could be made for MOTU_sample_1, which was over 99.5% identical to the SSU from *Helicotylenchus dihystera*, a plant ectoparasite. MOTU_sample_11 was nearly identical to the SSU from *Plectus aquatilis*, a free-living microbivore, and MOTU_sample_13 was identical to *Aporcelaimellus obtusicaudatus*, a predatory nematode. Using the extensive database of nematode SSU sequences (currently containing over 200 sequences from named taxa) other MOTU could be assigned to genera, as they cluster within known generic SSU diversity. Thus MOTU_culture_4, MOTU_culture_5 and MOTU_sample_15 were likely to be panagrolaims closely related to *Panagrolaimus* sp., a microbivore, and MOTU_sample_19 was likely to be an entomopathogenic steinernematid.

Morphological identification of the cultured isolates is congruent with the allocation of MOTU to named groups by cluster analysis. There remain some problems of resolution. The most abundant MOTU, observed 16 times in the random sample and 132 times in culture had sequences that differ by less than three bases from both *Cephalobus* and *Acrobeloides* species. These two genera are among the most confusing cephalobids even to the experts in the field. The diagnosis and separation of these genera is based currently on overlapping or loosely defined morphological characteristics and as a result it is difficult to put

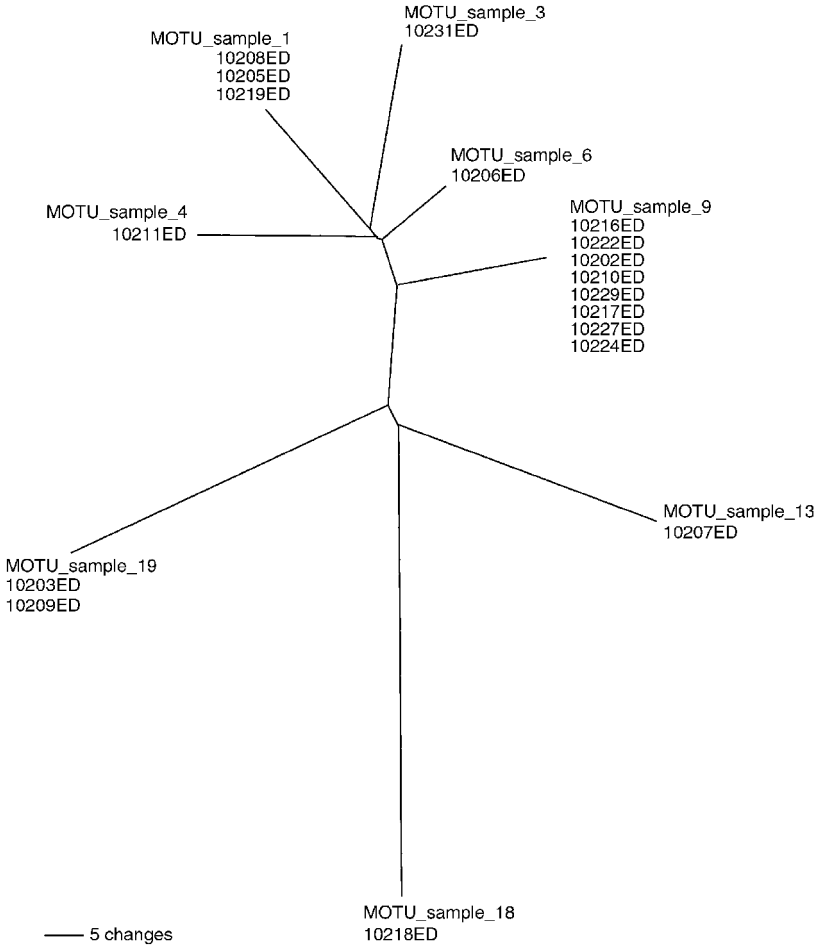

Fig. 3 Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from a random sample from a single subplot. Eighteen sequences from a single subplot of the Sourhope field site (designated 4 DU) were aligned and analysed as described in materials and methods. The analysis included 396 of the aligned nucleotides. The resultant tree is here represented as an unrooted phylogram, with branch lengths corresponding to those estimated from the uncorrected neighbour joining analysis (with missing and gapped sites excluded). Each cluster of sequences, identified by their specimen code, is designated with a MOTU number.

populations under one of the two based solely on morphological characteristics without a degree of uncertainty, indicating the inadequacy of morphology alone for their separation (De Ley *et al*. 1999). Both genera would be placed in the same MOTU by the heuristics employed here. Thus the methods are congruent, though in this case the MOTU approach does not distinguish the genera *Cephalobus* and *Acrobeloides*. It can be seen from Fig. 4 that the major nematode groups differ in the degree to which variation in SSU sequence correlates with morphologically based classification. Within the Cephalobidae, taxa classified as different genera (such as *Cephalobus*, *Acrobeloides* and *Cervidellus*) have similar or identical SSU sequences, while in the Rhabditidae, species within one genus (such as *Caenorhabditis elegans* and *C. briggsae*) have distinguishable sequences.

Light microscopic analysis of the five *Panagrolaimus* cultures revealed no morphological difference. Based on morphometry, however, the five cultures were categorized into two morphological groups, a large species (ED2021, ED2041, ED2042 and ED2043) and a small species (ED2013). Nevertheless, though culture ED2042 was closer in most measurements and de Man's ratios to the larger than to the smaller species, the fact that some measurements and ratios of culture ED2042 were intermediate is noteworthy (data not shown). The use of morphometry alone for the identification of *Panagrolaimus* has been criticised by Williams (Williams 1986) due to intraspecific variation (Mianowska 1977). Species that include both large and small individuals have been described (Borstom 1995) implying that size may not be an important identifying character within the genus. In this context, all five *Panagrolaimus* cultures could belong to the same morphological taxonomic unit, but can be separated into two groups on the basis of MOTU status.

Discussion

By sequencing an informative segment of DNA from a biological specimen it is possible to define molecular operational taxonomic units. To be useful, the segment of DNA must be known to be orthologous between species (as paralogues will define gene rather than organismal groups), and the segment must encompass sufficient variability to allow discrimination between MOTU useful to the research program. MOTU are identified through sequence identity. Identity in sequence need not correspond

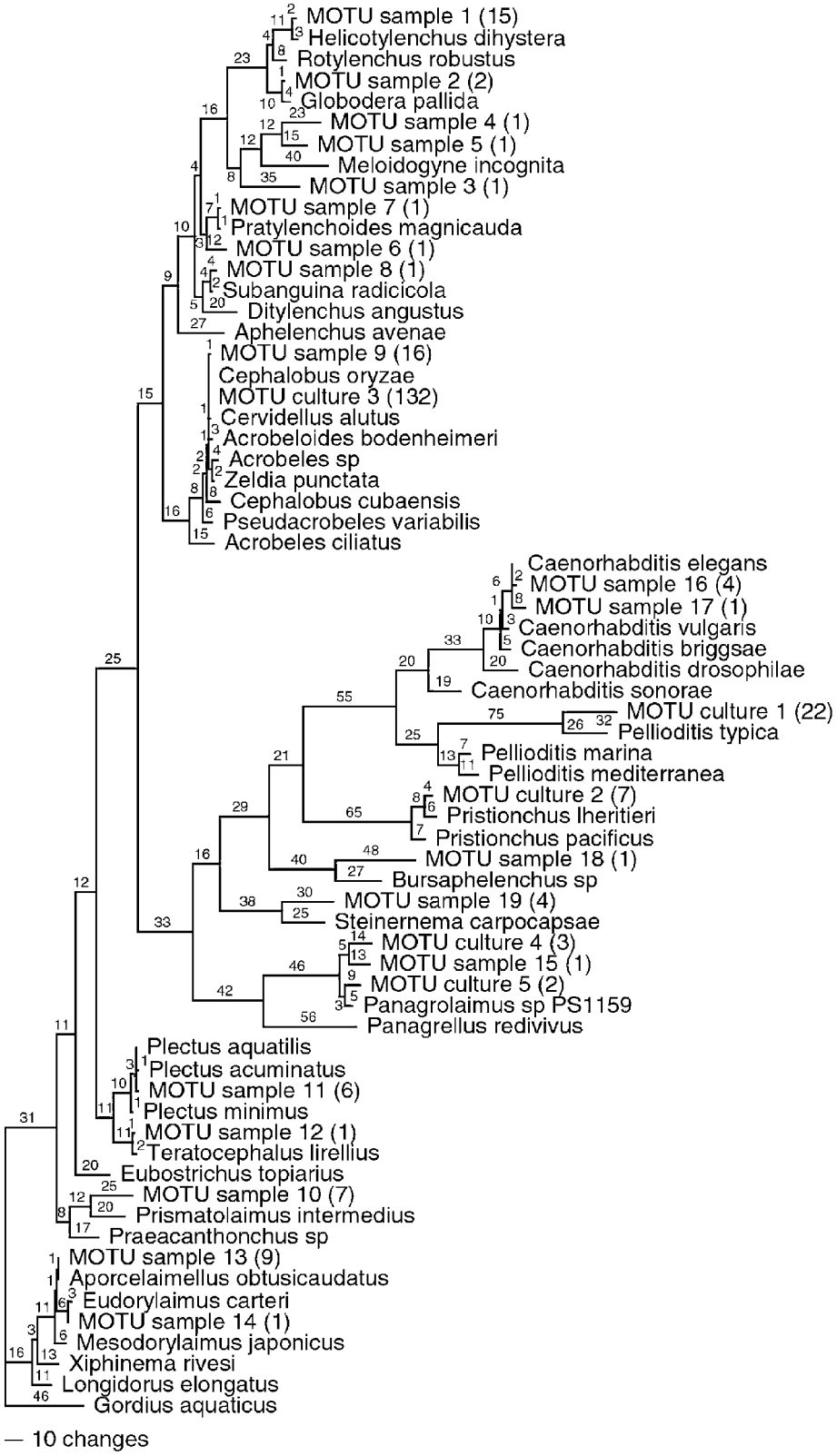

Fig. 4 Phylogram, rooted using the nematomorph *Gordius aquaticus* as an outgroup, of a neighbour joining analysis of all 24 survey sequences from this study and a selected set of 43 sequences from identified taxa. The alignment included 554 characters. The alignment was subjected to NJ analysis using the Kimura two-parameter distance correction. Branch lengths are given (in numbers of base changes). The MOTU are designated as in Figs 1–3, with a number in brackets indicating the number of sequences each represents.

to identity of operational taxonomic units (OTU) as measured by other models (biological or morphological): identity in sequence could mean 'the same taxon' or 'there is insufficient variation to define distinct taxa'. The same operational problem plagues other (biological or morphological) methods of defining taxa.

Differences in barcode sequence between specimens can arise in three ways. The differences might be part of the natural, within-OTU variation. Alternately, the differences could be due to methodological (sequencing) errors. These two types of difference should be disregarded when defining OTU. A third possibility is that the differences are related to a useful distinction between taxa. It is thus necessary (as with other methods, biological or morphological) to use heuristics for MOTU distinction based on known error rates in measurement, and perceived levels of difference that distinguish 'useful' MOTU. Importantly, for MOTU, unlike many OTU designators, these measures can be made explicit. For example, from known, accepted taxa within a particular group, the level of between-taxa within-group variation can be measured. Multiple re-sequencing of a single taxon will yield an observational error rate. The comparison between the between-taxon difference rate and the within-taxon variation and error rates will define the accuracy and specificity of the MOTU measurement. Given that it is clear from many gene sequences that different higher taxonomic groups can differ markedly in their background and adaptive substitution rates, and that different sized populations are expected to harbour different levels of within-taxon variation (also dependent on the population's evolutionary history), it may be necessary to define different heuristics for MOTU designation depending on the higher taxon studied.

The benefits of the MOTU approach are that data can be obtained from single specimens, often without compromising parallel or subsequent morphological identification [images of individuals can be recorded prior to PCR, or an individual can be dissected so that morphologically informative parts can be preserved while uninformative parts can be taken for PCR (Thomas *et al*. 1997)]; that morphologically indistinguishable taxa can be separated without the need for live material; and that a single technique is applicable to all taxa. Our extraction method also permits multiple PCR/sequencing events from a single specimen. Thus a long and partial training in morphological identification of a particular (sub) group is not necessary. All stages/morphs of taxa are amenable to study, as the method depends on genotype, not phenotype. In addition, the MOTU data, the sequences, are suited to exhaustive and model-driven phylogenetic analyses to derive independent and testable hypotheses of OTU interrelatedness.

We have here tested the 5′ end of the small subunit ribosomal RNA (SSU) gene as a MOTU identifier for soil nematodes. The pattern of conservation of SSU genes has made it possible to use it for both deep (interphylum and interkingdom) and local (generic) phylogenetic analyses. Analysis of available nematode full-length SSU sequences suggested that the SSU might be a good candidate for MOTU designation, as in many cases even closely related taxa were shown to have differences in their SSU sequence (Blaxter *et al*. 1998). SSU genes are commonly arranged as tandem arrays, for example *Caenorhabditis elegans* has one array of ~55 copies (Ellis *et al*. 1986; The *C. elegans* Genome Sequencing Consortium 1998). While genomic organization data is lacking for most nematodes, the similarity in organization of the known nematode SSU arrays (Sim *et al*. 1987; The *C. elegans* Genome Sequencing Consortium 1998) with those of other metazoans suggests that this pattern will be true of all nematodes. The repetitive nature of the SSU array makes it an easier target for PCR amplification, but also raises the problem of divergence between copies within an array. It is generally accepted that gene conversion and concerted evolution will tend to keep members of repeated gene arrays identical in sequence (Hillis & Dixon 1991), and there is no evidence in nematodes of one species carrying more than one very distinct SSU gene sequence variant. Prof. D. Fitch (personal communication) has been able to identify single base polymorphisms in nematode SSU genes: such variation would be classed within the same MOTU in our analysis.

Sequence similarity analyses, using the public databases (EMBL or GenBank) or a custom database of nematode small subunit ribosomal RNA sequences, of the MOTU barcode sequences allows identification of the individual nematodes as closely related to sequences derived from named taxa. These named-taxon sequences can be used to allocate the nematodes to known free-living, entomopathogenic and plant parasitic taxa. In the best case, there will be an exact match, and the MOTU can (provisionally) be allocated to a named taxon, and the biological attributes of that taxon can be transferred to the MOTU. In our dataset, we have many isolates from random sampling of a nematode SSU identical to that of the dorylaimid predator *Aporcelaimellus obtusicaudatus*. The reduced sequence similarity to other related Dorylaimidae (such as *Eudorylaimus carteri* and *Mesodorylaimus japonicus*, included in Fig. 4) suggests that this MOTU is likely to be in at least the same genus as *A. obtusicaudatus*, if not the same species. In support of this suggestion, we have also identified fixed specimens from Sourhope as *A. obtusicaudatus* (data not shown).

As the number of SSU sequences from identified nematodes is relatively small compared to the known or expected diversity of the phylum, such an exact match may be relatively uncommon, but the frequency of such matches will increase as additional SSU sequences are obtained and deposited in the public databases. However, using the molecular phylogenetic framework developed for the Nematoda, nonidentity can also be used to allocate

MOTU to genus or family level in taxonomic classifications. Such attributions can be made for all our MOTU. The attributions can aid in morphological identification of cultured specimens, by indicating which part of the diversity of nematodes they derive from. These allocations can also be employed to use the MOTU for ecological analyses, as biological features such as feeding mode and reproductive capacity can be inferred by comparison with known taxa. MOTU surveys can thus be used in overall diversity, ecological and other indices as would morphologically defined specimens.

Using 'nematode-universal' amplification primers, we were able to obtain PCR fragments and sequence from nematodes that map across the wide range of diversity in the Phylum Nematoda. Barcode sequences were obtained for taxa in Clades I, II, IV and V as defined by molecular phylogenetic analysis (note that clade III is exclusively animal-parasitic) (Blaxter *et al*. 1998). The method thus appears applicable to all nematodes, and not restricted to a specific phylogenetic group. There was no apparent correlation between stage and size of nematode and the success of the technique. We thus believe that we are not systematically missing aspects of the diversity. Our current MOTU diversity from the random survey at Sourhope is 23 taxa. This value is derived from only 240 sequences. We cannot yet robustly estimate the total number of taxa to be defined by MOTU at Sourhope, but the result of resampling a single subplot independently, as illustrated in Fig. 3, suggests that we are currently some way from saturating our sampling of the site: the 'collector's curve' is still on the rise. Intensive sampling of grassland ecosystems has been carried out at Kansas (USA), Porton Down (UK) and several sites in Eastern Europe (Austria, Poland, Romania, Slovakia). The highest number of (morphologically identified) species is in Kansas, where 228 taxa are recorded (Orr & Dickerson 1966; Boag & Yeates 1998). In the UK, the maximum number recorded is at Porton Down, where a chalk grassland yielded 154 taxa (Hodda & Wanless 1994). Overall, Boag and Yeates calculated the mean published species diversity in grasslands to be 42.8 taxa (with a range from 6 to 228) (Boag & Yeates 1998). In terms of upland grass ecosystems dominated by *Festuca* species, 18–27 species have been recorded in single survey samples (Yeates 1974). We have compared the taxonomic distribution and abundances of major taxonomic groups identified by the MOTU method at Sourhope and the relevant morphological surveys of Hodda & Wanless (1994) and Yeates (1974) (Table 2). The Sourhope soil nematode fauna has a similar distribution in terms of numerical abundance to the other sites, particularly the New Zealand Cluden site, a *Festuca* grassland (Yeates 1974). The Sourhope site is relatively abundant in chromadorids, represented by a MOTU most closely related to *Prismatolaimus*, as would be expected from the climatic and soil conditions (high precipitation and water retention). We note that the taxon distribution per major taxonomic group is more disparate between the MOTU survey of Sourhope and the morphological surveys, particularly in an under-representation of dorylaimid taxa. This may have resulted from the small size of our sample thus far, and should be rectified by more exhaustive sampling now underway. However, another possibility is that our MOTU approach is of insufficient resolution to distinguish nematodes in this group. There may be taxa present which morphologists would recognize as distinct, but which have little or no variability in the SSU fragment sequenced here, and thus fall into the same MOTU. There are relatively few

Table 2 Comparison of MOTU method applied to Sourhope and other grassland nematode surveys by major taxonomic group[1]

Order	No. of taxa	% of taxa	Abundance	% of abundance	No. of taxa	% of taxa	Abundance	% of abundance
	A	**Yeates — Conroy**			**B**	**Yeates-Cluden**		
Tylenchida	6	37.50	172	58.31	8	34.78	110	36.91
Rhabditida	4	25.00	68	23.05	3	13.04	101	33.89
Areolaimida	1	6.25	9	3.05	2	8.70	41	13.76
Monhysterida	0	0.00	0	0.00	1	4.35	3	1.01
Chromadorida	0	0.00	0	0.00	1	4.35	1	0.34
Dorylaimida	5	31.25	46	15.59	8	34.78	42	14.09
Total	16		295		23		298	
	C	**Hodda & Wanless — Porton Down**			**D**	**Sourhope**		
Tylenchida	63	39.62	2640	58.46	9	47.37	24	32.43
Rhabditida	27	16.98	701	15.52	5	26.32	26	35.14
Areolaimida	18	11.32	261	5.78	2	10.53	7	9.46
Monhysterida	2	1.26	91	2.02	0	0.00	0	0.00
Chromadorida	6	3.77	76	1.68	1	5.26	7	9.46
Dorylaimida	43	27.04	747	16.54	2	10.53	10	13.51
Total	159		4516		19		74	

[1]Data are taken from A,B: Yeates (1974); C: Hodda & Wanless (1994) and D: this study (excluding culture-only MOTUs).

sequences available from Dorylaimidae to test the within-family variability, but sequences from *Eudorylaimus carteri* and *Mesodorylaimus japonicus* have been included in our analysis (Fig. 4), and are distinguished. This suggests that our method is able to resolve taxa at least at the genus level in this family. However, a parallel morphological survey will be needed to determine in detail how the diversity measured by molecular methods correlates with that found by traditional classification, and we plan to carry out such a survey at a later date.

These initial results using the SSU MOTU technique are, in our view, very promising. We are continuing to sample soil nematodes from the Sourhope field site using the system outlined herein, with modifications to increase throughput. In particular we are automating the base calling, sequence trimming, alignment and phylogenetic analysis steps. Several sequences were excluded from analysis because of overall low base quality calls and the sequencing step is also being optimized. We are also testing alternative methods of nematode extraction, since the paper filtration method used for our intial survey may have introduced some bias into our sampling. We are investigating the relationship between MOTU and 'biological' species by correlating the morphological allocation of cultured nematodes to species, their ability to interbreed, and MOTU. We are building a larger database of SSU barcodes from random samples from Sourhope, and other sediments, including littoral and marine nematodes. The approach we have taken to build the database of diversity using MOTU, is relatively expensive in terms of consumables, though very efficient in time. For more extensive surveys, a cheaper, oligonucleotide-hybridization approach could be taken, where the SSU PCR products are arrayed on filters or microarrayed on slides and identified by probes derived from diagnostic SSU fragments from known or indicator taxa, chosen for their relevance to the study in question.

Acknowledgements

We would like to thank the other members of the Blaxter lab and the Sourhope field staff, in particular Dr Sarah Buckland, for support, and Jill Lovell for technical assistance with sequencing. Dr David Fitch, Dr Tom Bongers, Vincent Scholze and Hanny van Megen aided in morphological identification of cultured nematodes. Dr Hans Helder and Dr Sven van den Elsen helped with discussion of molecular methods. The project was initially conceived with significant input from Dr Armand Leroi, and has benefited from the expertise and enthusiasm of Dr Paul De Ley. Dr Josephine Pemberton gave helpful comments on the manuscript. This work was funded by the Natural Environment Research Council Soil Biodiversity and Ecosystem Function Programme.

References

Adams BJ (1998) Species concepts and the evolutionary paradigm in modern nematology. *Journal of Nematology*, **30**, 1–21.

Adams BJ, Burnell AM, Powers TO (1998) A phylogenetic analysis of the genus *Heterorhabditis* (Nemata: Rhabditidae) based on internal transcribed spacer 1 DNA sequence data. *Journal of Nematology*, **30**, 22–39.

Beckenbach K, Blaxter ML, Webster JM (2000) Phylogeny of *Bursaphelenchus* species derived from analysis of ribosomal internal transcribed spacer DNA sequences. *Nematology*, **1**, 539–548.

Bird DM, Riddle DL (1994) A genetic nomenclature for parasitic nematodes. *Journal of Nematology*, **26**, 138–143.

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

Boag B, Yeates GW (1998) Soil nematode biodiversity in terrestrial ecosystems. *Biodiversity and Conservation*, **7**, 617–630.

Borstom S (1995) Populations of *Plectus acuminatus* Bastian, 1865 and *Panagrolaimus magnivulvatus* n. sp. (Nematoda) from nunatakks in Dronning Maud Land, East Antarctica. *Fundamental and Applied Nematology*, **18**, 25–34.

Cobb NA (1918) Estimating the nema population of soil, with special reference to the sugar beet and root gall nemas, *Heterodera schachtii* Schmidt and H. Radicicola (Greef) Muller, and with a Description of *Tylencholaimus Aequalis n.sp*. In: Agricultural Technology Circular. Bureau of Plant Industry. U.S. Department of Agriculture, Washington, DC, USA.

De Grisse AT (1969) Redescription ou modifications de quelques techniques utilisées dans l'étude des nematodes phytoparasitaires. *Mededelingen Rijksfakulteit Landbouwwetenschappen, Gent*, **34**, 351–369.

De Ley P, Blaxter ML (2001) Systematic position and phylogeny. In: *The Biology of Nematodes* (ed. Lee D). Harwood Academic Publishers, Reading.

De Ley P, Felix M-A, Frisse LM, Nadler SA, Sternberg PW, Thomas WK (1999) Molecular and morphological characterisation of two reproductively isolated species with mirror-image anatomy (Nematoda: Cephalobidae). *Nematology*, **1**, 591–612.

Dorris M, De Ley P, Blaxter M (1999) Molecular analysis of nematode diversity. *Parasitology Today*, **15**, 188–193.

Ellis RE, Sulston JE, Coulson AR (1986) The rDNA of *C. elegans*: sequence and structure. *Nucleic Acids Research*, **14**, 2345–2364.

Elson-Riggins JG, Al-Banna L, Platzer EG, Kaloshian I (2001) Characterization of *Otostrongylus circumlitus* from Pacific harbor and northern elephant seals. *Journal of Parasitology*, **87**, 73–78.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Folkertsma RT, Rouppe van der Voort JN, de Groot KE *et al.* (1996) Gene pool similarities of potato cyst nematode populations assessed by AFLP analysis. *Molecular Plant—Microbe Interactions*, **9**, 47–54.

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, **66**, 411–436.

Hodda M, Wanless FR (1994) Nematodes from an English chalk grassland: species distributions. *Nematologica*, **40**, 116–132.

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, **180**, 4765–4774.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences*, **23**, 403–405.

Lambshead J (1993) Recent developments in marine benthic biodiversity research. *Oceanis*, **19**, 5–24.

Lambshead PDJ (2001) Marine Nematode Diversity. In: Nematology, advances and perspectives. (eds Chen ZX, Chen SY, Dickson DW). ACSE-TUP Book Series, San Francisco, USA.

Lawton JH, Bignell DE, Bolton B *et al.* (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72–75.

Malakhov VV (1994) *Nematodes. Structure, Development, Classification and Phylogeny*. Smithsonian Institution Press, Washington.

Mianowska E (1977) Research on the biology and ecology of *Panagrolaimus rigidus* (Schneider) (Thorne). VI. The influence of the population's origin and breeding conditions on morphometric features. *Ekolgia Polska*, **25**, 323–331.

Orr CC, Dickerson OJ (1966) Nematodes in true prairie soils of Kansas. *Kansas Academy of Sciences Transactions*, **69**, 317–334.

Platt HM (1994) Foreword. In: *The Phylogenetic Systematics of Free-Living Nematodes* (ed. Lorenzen S). The Ray Society, London.

Powers TO (1992) Molecular diagnostics for plant nematodes. *Parastology Today*, **8**, 177–179.

Powers TO, Adams BJ (1994) Nucleotide sequences in nematode systematics. In: *Advances in Molecular Plant Nematology* (ed. Lamberti F), pp. 99–108. Plenum Press, New York.

Powers TO, Harris TS (1993) A polymerase chain reaction method for identification of five major *Meloidogyne* species. *Journal of Nematology*, **25**, 1–6.

Powers TO, Todd TC, Burnell AM *et al.* (1997) The internal transcribed spacer region as a taxonomic marker for nematodes. *Journal of Nematology*, **29**, 441–450.

Seinhorst JW (1959) A rapid method for the transfer of nematodes from fixative to anhydrous glycerine. *Nematologica*, **4**, 67–69.

Semblat JP, Bongiovanni M, Wajnberg E *et al.* (2000) Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne* spp. *Heredity*, **84**, 81–89.

Semblat JP, Wajnberg E, Dalmasso A, Abad P, Castagnone-Sereno P (1998) High-resolution DNA fingerprinting of parthenogenetic root-knot nematodes using AFLP analysis. *Molecular Ecology*, **7**, 119–125.

Sim BKL, Shah J, Wirth DF, Piessens WF (1987) Characterisation of the filarial genome. In: *Filariasis (Ciba Foundation Symposium 127* eds. Evered D, Clark S), pp. 107–124. Wiley, Chichester (UK).

Southey, ed. (1986) *Laboratory Methods for Work with Plant and Soil Nematodes*. Reference Book. Ministry of Agriculture, Fisheries and Food. Her Majesty's Stationery Office, London.

Stanton JM, McNicol CD, Steele V (1998) Non-manual lysis of second stage *Meloidogyne* juveniles for identification of pure and mixed samples based on polymerase chain reaction. *Australian Plant Pathology*, **27**, 112–115.

Swofford D (1999) *PAUP* 4.0b6*. Sinauer Associates, Sunderland, MA, USA.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic Inference. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 407–514. Sinauer Associates, Sunderland, MA, USA.

Szalanski AL, Sui DD, Harris TS, Powers TO (1997) Identification of cyst nematodes of agronomic and regulatory concern by PCR-RFLP of ITS1. *Journal of Nematology*, **29**, 253–264.

The C. elegans Genome Sequencing Consortium (1998) Genome sequence of *Caenorhabditis elegans*: a platform for investigating biology. *Science*, **282**, 2012–2018.

Thomas WK, Vida JT, Frisse LM, Mundo M, Baldwin J (1997) DNA sequences from formalin-fixed nematodes: integrating molecular and morphological approaches to taxonomy. *Journal of Nematology*, **29**, 248–252.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Williams MSR (1986) The use of scanning electron microscopy in the taxonomy of *Panagroliamus* (Nematoda: Panagrolaimidae). *Nematologica*, **32**, 89–97.

Woese CR (1996) Phylogenetic trees: Whither Microbiology? *Current Biology*, **6**, 1060–1063.

Yeates GW (1974) Studies on a climosequence of soils in tussock grasslands. 2. Nematodes. *New Zealand Journal of Zoology*, **1**, 171–177.

Zhou J, Davey ME, Figueras JB *et al.* (1997) Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. *Microbiology*, **143**, 3913–3919.

The Blaxter lab at the Institute of Cell, Animal and Population Biology conducts studies on the phylogenetics, genetics and genomics of nematodes, both parasitic and free living. Details may be obtained from the website http://www.nematodes.org/ .Nucleotide sequences reported in this paper have been deposited in the public databases with accession numbers AF430402-AF430641; The alignments used have been deposited in the EMBL database with accession numbers ALIGN_000248-ALIGN_000249.
