# Molecular Assessment of Soil Nematode Diversity

Robin M. Floyd

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Edinburgh

2003

I declare that this thesis has been composed by myself and that, except where stated otherwise, all work presented here is my own.

# Contents

# Abstract

Free-living nematodes are an example of a group of organisms which presents great challenges to traditional taxonomy – they are highly species-rich, numerically abundant and present in virtually all soil and marine sediment habitats. In order to address vital questions concerning the links between biodiversity and ecosystem function, it is first necessary to have a means of measuring diversity: of defining taxa, of quantifying their relative abundances, and of mapping their distributions. This work presents a novel set of methods allowing measurement of nematode diversity through DNA sequence. A survey was carried out at a grassland field site in the Scottish Southern Uplands, from which a set of individual nematode DNA sequences was generated (of the small subunit ribosomal RNA gene, or SSU). Specimens were clustered into molecular operational taxonomic units (MOTU) based on pairwise comparisons of sequence identity. MOTU were assigned to traditional taxonomic groups by comparison to a set of SSU sequences from known nematode taxa. Various aspects of the method were tested for robustness, in particular the effect of varying sequence processing order, and of altering the level of sequence identity taken to define MOTU. The sequencing error rate was also estimated, and utilised to distinguish between sequence differences due to real variation, and those due to experimental errors. Various diversity indices and other measures of biological diversity were explored. These measures were used to test for correlations between diversity and environmental factors, including a set of experimental treatments applied at the field site. These results suggest that molecular survey methods provide a powerful and effective means of analysing patterns in diversity within groups of organisms such as nematodes.

# 1. Introduction

To any observer, the most prominent feature of life is surely its diversity. In numbers and in varieties, living things stretch beyond our capacity to comprehend. Yet the earth's biota is being altered at an unprecedented rate by human activity (Schulze and Mooney 1994), with unknown consequences. We do not know to within an order of magnitude how many species of living organism exist on the planet (May 1988); still less do we understand how those organisms give rise to the ecosystem processes upon which all life depends (Schwartz et al. 2000). We will be unable to comprehend or predict the results of changes to natural ecosystems until we have a means of measuring the diversity of life that exists. For, although large and conspicuous animals such as birds and mammals are familiar, well-sampled and thoroughly-classified, we have only begun to appreciate the numbers and variety of those organisms which thrive just beyond our usual notice.

Free-living nematodes are an example of such a group of organisms. Ubiquitous in soil and marine sediments, they are among the most abundant and yet poorly characterised animal taxa. They may often be found at densities of 1 million individuals per square metre; thus the nematodes in a typical hectare of soil far outnumber all the human beings in the world. Several thousand species have been formally described in the scientific literature (Malakhov 1994), but it is evident that this is a significant underestimate of the total number in existence, as every thorough survey of a new environment has uncovered yet more undescribed species (Lawton et al. 1998; Lambshead 2001). Estimates of the total number range from 80,000 (Malakhov 1994) to 100,000,000 (Lambshead 1993). The number of undescribed nematode taxa on earth is far in excess of the number of nematologists, yet we have only recently begun to understand the many important ecological roles played by nematodes: as predators and prey of other soil organisms, as secondary decomposers, as determinants of plant productivity (Malakhov 1994), and as "bioindicators" of environmental disturbance and pollution (Bongers, 1990).

The relative neglect of nematodes in ecological surveys thus far has been largely due to the practical difficulties associated with their analysis: their microscopic size and lack of easily distinguishable morphological features makes identification difficult, with considerable training and expertise necessary, while their sheer abundance means that an exhaustive survey of even a single habitat is likely to be an impossible task (Lawton et al. 1998). What is sorely needed is a means of surveying nematode diversity that is robust, time-efficient, and universally applicable.

## 1.1 Biodiversity: kinds and numbers

The study of the variety of living organisms in nature is the field of biodiversity (a contraction of "biological diversity"). The US Congress Office of Technology Assessment offers the following definition of this term:

> "Biological diversity is the variety and variability among living organisms and the ecological complexes in which they occur. Diversity can be defined as the number of different items and their relative frequency. For biological diversity, these items are organised at many levels, ranging from complete ecosystems to the chemical structures that are the molecular basis of heredity. Thus, the term encompasses different ecosystems, species, genes and their relative abundance." (OTA 1987).

Therefore, in order to make biodiversity a measurable quantity rather than an abstract concept, a basic requirement is to define measurable entities, whether they are genes, individuals, species or other taxonomic categories. Biodiversity, then, may be measured in terms of kinds, numbers and their distributions. But what 'kinds' are appropriate to measure? The human mind has always approached the world by placing objects and phenomena in categories. Some categories represent real patterns in nature, while others merely reflect human perceptive and cognitive biases (Hey 2001). In science, we should always seek to discover the former and reject the latter. It has always been intuitively clear that living things exists as distinct and recognisable 'kinds' of organisms, but also that individual organisms within a 'kind' vary from one to another.

The most commonly used measure of life's diversity remains the species. In the taxonomic system of Linnaeus, still in use today, the species is the lowest level, and might therefore be considered the most basic unit of diversity – since genera are merely groupings of species, families groupings of genera, and so on. The concept of "biodiversity hotspots" (Myers et al. 2000), for example, has been proposed as a basis for conservation policy. Hotspots are "areas featuring exceptional concentrations of endemic species and experiencing exceptional loss of habitat" - a definition stated explicitly in terms of species. Indeed, species have been referred to as "the units of biodiversity" (Claridge et al. 1997). Since the term is used across all kingdoms of life, it might be expected to mean the same thing wherever it is used. But how valid is this assumption? For example, a biodiversity survey of a given habitat might record 200 nematode species and 100 insect species. But if an insect species is defined by different criteria than a nematode species, how meaningful is this comparison? And how certain are we that we are recovering objective, meaningful units of biodiversity rather than mere arbitrary categories? In fact, "species" itself is a term surrounded by much controversy, for which there is no universally agreed-upon definition.

## 1.2 Species concepts

The species was established by Linnaeus as the lowest level of the taxonomic hierarchy. In this system, species were considered to represent the essential created kinds of organisms. Within species, variation was possible, but between them fixed boundaries existed - one species could never change into another. Species were defined on the basis of morphological similarity; the defining features obviously differed in every group of organisms, and the only criterion for deciding which were important was the subjective judgement of taxonomists.

This view was overturned by Darwin, who showed that there were no immutable kinds, and that current forms of life had evolved from different ones in the past (Darwin 1859). Lineages which once represented mere varieties of the same species could diverge, through gradual and continuous change over long periods of time, into different species; by the same reasoning, species could diverge into genera, and all other higher taxonomic categories represented divergences in the increasingly distant past. The difference between varieties, species and higher categories, then, was only a matter of degree. Indeed, Darwin did not consider species to be a fundamentally important category: "I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms. The term variety, again, in comparison with mere individual differences, is also applied arbitrarily, and for mere convenience sake." (Darwin 1859; p42 in the 6th edition).

Although Darwin's theory of evolution became the foundation of biology, it seems that most authors following him, including most modern authors (e.g. Eldredge and Cracraft 1980; Claridge et al. 1997; Futuyma 1998) maintained the view that species have an inherent biological reality. Therefore, there have been many attempts to establish universal, consistent definitions by which we may recognise true species in nature.

Perhaps the first of these was the biological species concept (BSC), expressed by Mayr (1963) as follows: "species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups". However, this concept has two major limitations (Noor 2002): first, it cannot be applied to asexual species; and second, it provides no clear answer to whether allopatric groups which would never meet under natural conditions should be considered separate species if they can nevertheless produce fertile offspring when artificially brought together (e.g. lions and tigers). It also includes no historical dimension - it only allows us to distinguish species existing at the same moment, and makes no reference to the fact that species are lineages continuing through time (Ridley 1996). A further criticism is that if we recognise species by reproductive isolation, and at the same time define the cause of speciation as reproductive isolation, this amounts to circular reasoning (Mallet 1995). Mallet and others (Bush 1994; Dover 1995) have argued that species cannot be defined in terms of an evolutionary process assumed to give rise to them, as this would prejudice us against other processes, such as sympatric speciation. Such problems with the BSC have led to a proliferation of alternatives. Mayden (1997;1999)

lists 22 species concepts which have appeared in the literature (though some of these may be considered essentially synonymous), based on such criteria as ecological niches, mate recognition, cohesion and evolutionary history.

The lack of any forthcoming resolution to this controversy is perhaps due to the conflict between two distinct goals of any species concept, which are not always explicitly separated. When we ask "what is a species?" we are implicitly asking two different questions. The first is a 'conceptual' question of what in reality is meant by the term species, if anything. The second is an 'operational' question about how, in practice, we go about recognising such entities in the real world. Satisfying both of these requirements at the same time is a significant problem. It may even be an irreconcilable one, as was argued by Adams 2001), who expressed this difficulty as the 'species delimitation uncertainty principle', analogous to the Heisenberg uncertainty principle of quantum mechanics. Applied to species concepts, this states that "attempts to make species concepts operational come at the expense of theoretical rigour (and vice versa)."

As an example, consider the concept with which we began, that of species as bifurcating evolutionary lineages. This is essentially the evolutionary species concept (ESC), defined as: "A species is a single lineage of ancestor-descendant populations which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate" (Wiley 1978). Many authors (e.g. Brooks and McLennan 1999; Mayden 1999; Wheeler 1999) consider that this concept is the closest to fundamentally defining what species are, but it is a non-operational concept, as it defines species in theoretical terms but does not specify how we should recognise such entities in practice. An example at the opposite extreme is the suggestion (Hagstrom et al. 2002), based on DNA sequence comparisons of marine bacterioplankton, that "a 16S rDNA sequence similarity of $\geq$97% is a reasonable level for grouping bacteria into species." This is clearly an operational concept, but conceptually it does not address the question of what species are, or why we should be able to recognise units that are in some way biologically significant on the basis of 97% 16S rDNA similarity, rather than 96% or 98%. Nor is it universally applicable – the definition as stated applies only to bacteria (or at least to organisms with 16S rDNA genes).

In between these two extremes, many authors have sought concepts which act as operational surrogates for the ESC, such as the phylogenetic species concept (PSC): that species are "the smallest aggregation of (sexual) populations or (asexual) lineages diagnosable by a unique combination of character states in comparable individuals" (Cracraft 1997). This concept has several advantages: it is testable and broadly applicable, and it emphasises that species are evolutionary lineages (Wheeler 1999). A potential disadvantage is that it may be possible to diagnose unique character states for every population, however trivial, and therefore, according to the PSC, each of these would have to be assigned species status. This could lead to an increase of orders of magnitude in the number of named species (Knowlton and Weight 1997). Mallet (1995) argued that species should be defined only in terms of the patterns observed in nature, rather than in terms of concepts assumed to give rise to them. He proposed that species are simply "genotypic clusters" with few or no intermediates.

Some authors (e.g. Hull 1997) have argued that there is no single species concept which we should expect to be applicable in every case - that is, that we should adopt a pluralistic view of species, incorporating several concepts. When one fails, we can switch to another. Along similar lines, Mayden (1977;1999) proposes a hierarchy of species concepts, with one, the evolutionary species concept, fundamental to all the rest.

Where does this controversy leave us in our urgent need to assess the biodiversity of the planet? Arguments over the species concept look set to continue for some time, and may even, for the reasons discussed above, be fundamentally irreconcilable; yet the 'biodiversity crisis' is real, and a universally agreed upon species concept may come too late to be useful, if ever. It could be argued that the constant focus on finding the one true concept of 'species' has actually held back biology in this respect, with scientists perhaps subconsciously holding on to a pre-Darwinian, 'essentialistic' or 'idealistic' notion of life as a collection of fixed entities. Perhaps we would better understand life by considering it as Darwin did, as a continuum of variation, regardless of where we choose to place the dividing lines. With this in mind, what is needed is a purely operational measure of biodiversity – a means of categorizing living things that is practical and applicable, without necessarily satisfying every theoretical consideration about precisely what the categories mean. This was the principle behind phenetic or numerical taxonomy (Sokal and Sneath 1963). In this approach, essentially any means of classifying organisms is considered valid, with the condition that every element of the methodology must be explicitly specified: the characters used, the clustering method applied, and the exact level of similarity taken to define taxa. The entities defined in this way are termed 'operational taxonomic units' (OTUs). This is a neutral term referring to a taxonomic grouping of any status; species, genera, families etc. are all specific cases of OTUs.

Here, then, is a potential solution to the problem of surveying the diversity of 'difficult' groups of organisms. OTUs could be defined which are relevant to the group of organisms and the study at hand, using characters which are easily measured without requiring great taxonomic expertise, and which should be consistently repeatable by different scientists across different study sites.

## 1.3 An Operational Taxonomy for Nematodes

Free-living nematodes would appear to be a prominent example of a group of organisms whose study could benefit from an OTU approach. If it were possible to sample and identify nematodes with a greater throughput and without expert training, many nematode communities could be better characterised. The question becomes which features are appropriate to use. It would be possible to use morphological characters, perhaps through establishing a universally agreed-upon character scoring scheme, but this approach retains many of the same problems as traditional classification – the microscopic size and lack of clearly distinguishable morphological features still means that every individual specimen must be subjected to time-consuming examination; the subjective nature of observations creates difficulties in ensuring

between-experiment and between-laboratory consistency in classification; and certain individuals of the 'wrong' sex or life-cycle stage may lack key identifying features.

In recent years, molecular biology has produced a wide variety of tools allowing us to examine organisms in new ways. Rather than defining taxa in terms of morphological traits (the phenotype), it is now possible to analyse the underlying DNA sequences (the genotype). A specified sequence (or other characteristic molecular pattern, such as restriction fragment length polymorphisms or randomly amplified polymorphic DNA) can be used to define a molecular operational taxonomic unit (MOTU) (Floyd et al. 2002). Such an approach brings a number of advantages: molecular markers can be described rigorously and without observer bias (two DNA sequences are either identical or they are not); conversely, there is an experimental error rate associated with sequencing, but this too can be measured and taken into account. As with any method of OTU designation, molecular or morphological, it is necessary to use heuristics based on known observational error rates, and on the level of variation perceived to correspond to 'useful' taxa. However, unlike many methods of OTU designation, with molecular methods these heuristics can be explicitly specified, and potentially varied for different analyses. Molecular methods are applicable to any individual, regardless of sex or life-cycle stage; there is, however, expected to be a certain 'fail' rate, which may be random with respect to taxon, or may be biased, which again can be determined by experimentation. Since molecular methods are amenable to automation and high-throughput processing, this approach allows rapid and efficient assessment of a community.

A molecular marker used for MOTU designation must fulfil a number of criteria. It must be known to be orthologous, not paralogous, between organisms compared; it must be sufficiently variable to discriminate taxa useful to the research program at hand, yet also sufficiently similar to allow use of universal primers for PCR and sequencing, and to permit alignment of disparate sequences for comparison. Ideally, it should be possible to place even an entirely novel sequence in relation to a known taxonomic group by comparison with existing sequence data from known groups.

A wide range of methods have been utilised in the analysis of molecular variation. The restriction fragment length polymorphism (RFLP) method exploits variation in the number and location of restriction enzyme target sites, so that when PCR products are digested with a particular enzyme, the resulting fragments of varying sizes create a unique DNA banding pattern on an elecrophoresis gel (Powers and Harris 1993; Powers et al. 1997; Szalanski et al. 1997). The amplified fragment length polymorphism (AFLP) technique is based on a restriction digest of total genomic DNA, followed by selective PCR amplification of sets of restriction fragments (Vos et al. 1995; Semblat et al. 1998). Randomly amplified polymorphic DNA (RAPD) is another whole-genome analysis method, in which arbitrary primers are used to generate a large number of random fragments whose identities are unknown, but overall will generate a unique pattern for a given taxon. RFLP, AFLP and RAPD are all methods which generate a unique "fingerprint"-like banding pattern for a given taxon and display a large amount of information (e.g. hundreds of fragments), making them useful for fine-level discrimination between species or populations within particular, known taxonomic groups, but the high level of variability and the complexity of the

information generated makes it unsuitable for a general-purpose taxonomic system, as there is no way for a novel pattern to be related to any known taxonomic group. Another method is denaturing gradient gel electrophoresis (DGGE), which allows separation of DNA molecules of similar length but different sequence, based on the fact that slight differences in sequence can often result in changes in the electrophoretic mobility of DNA fragments within a linear gradient of denaturants (Foucher and Wilson 2002). Thus a mixed population of PCR products can be separated into a number of distinct fragments as bands on a gel. This provides a simple and inexpensive means of determining the broad taxonomic diversities of different samples. However, the possibility remains that distinct sequences may have identical denaturing behaviour and thus will not be distinguishable by this method; also, no information is provided which would allow particular bands to be unequivocally related to known taxa.

Finally, we may directly sequence the genomic DNA of the organisms of interest. A taxon might be considered to be best described by its entire genome, but it is not possible in practice to sequence the whole genome of every individual organism we wish to indentify, any more than it is feasible to record every possible morphological trait that exists. However, just as taxonomists have always worked by picking a set of informative morphological characters for analysis, a particular segment of DNA could be taken to stand for the whole genome, or at least as representative of variation between genomes.

For this study, the sequence of the 5' end of the nuclear small subunit ribosomal RNA gene (SSU or 18S) has been chosen as the molecular marker for designating soil nematode MOTU. This is a part of the ribosomal DNA (rDNA) cluster (see Figure 1.1), a multi-copy repeating unit containing the genes which encode the RNA components of the ribosomes. The function of the ribosomal RNA (rRNA) encoded by the SSU gene is dependent on the formation of a secondary structure containing both single-stranded loop and double-stranded stem regions; the sequence within the stem regions is evolutionarily constrained by the necessity to form the correct structure with the opposite strand, while the loop regions are free to vary by mutation as, in most cases, changes in sequence have no impact on the product molecule's function. This property of containing both conserved and variable regions within the same gene makes the SSU - as well as other rRNA-encoding genes such as the large subunit (LSU) - useful for phylogenetic analysis and taxon discrimination at a variety of levels, from deep (e.g. inter-phylum) to local (e.g. specific and generic) (Hillis and Dixon 1991). Ribosomal genes are typically found in tandem arrays within eukaryotic nuclear genomes; the nematode *Caenorhabditis elegans*, for example, has an array of around 55 copies (The *C. elegans* Genome Sequencing Consortium 1998). While this large copy number facilitates PCR amplification of rDNA genes, it also creates the potential for divergence between copies within a single genome. However, there is evidence that members of the same array will tend to remain identical, or nearly identical, due to concerted evolution (Arnheim et al. 1980; Hillis and Dixon 1991) largely eliminating this problem.

As a result of its phylogenetic utility, a large dataset of SSU sequences from known nematode taxa has already been generated (Nadler 1992; Zarlenga et al. 1994; Fitch et al. 1995; Aleshin et al. 1998a; Aleshin et al. 1998b; Blaxter et al. 1998; Kampfer et al. 1998; Dorris et al. 1999; Felix et al. 2000), and is

available in public databases. This means that, in a molecular survey, a sequence derived from an unidentified nematode may often be matched to a known sequence, and hence to a known taxon; or, if a sequence is novel, phylogenetic methods should allow its placement within the known tree of nematode diversity.

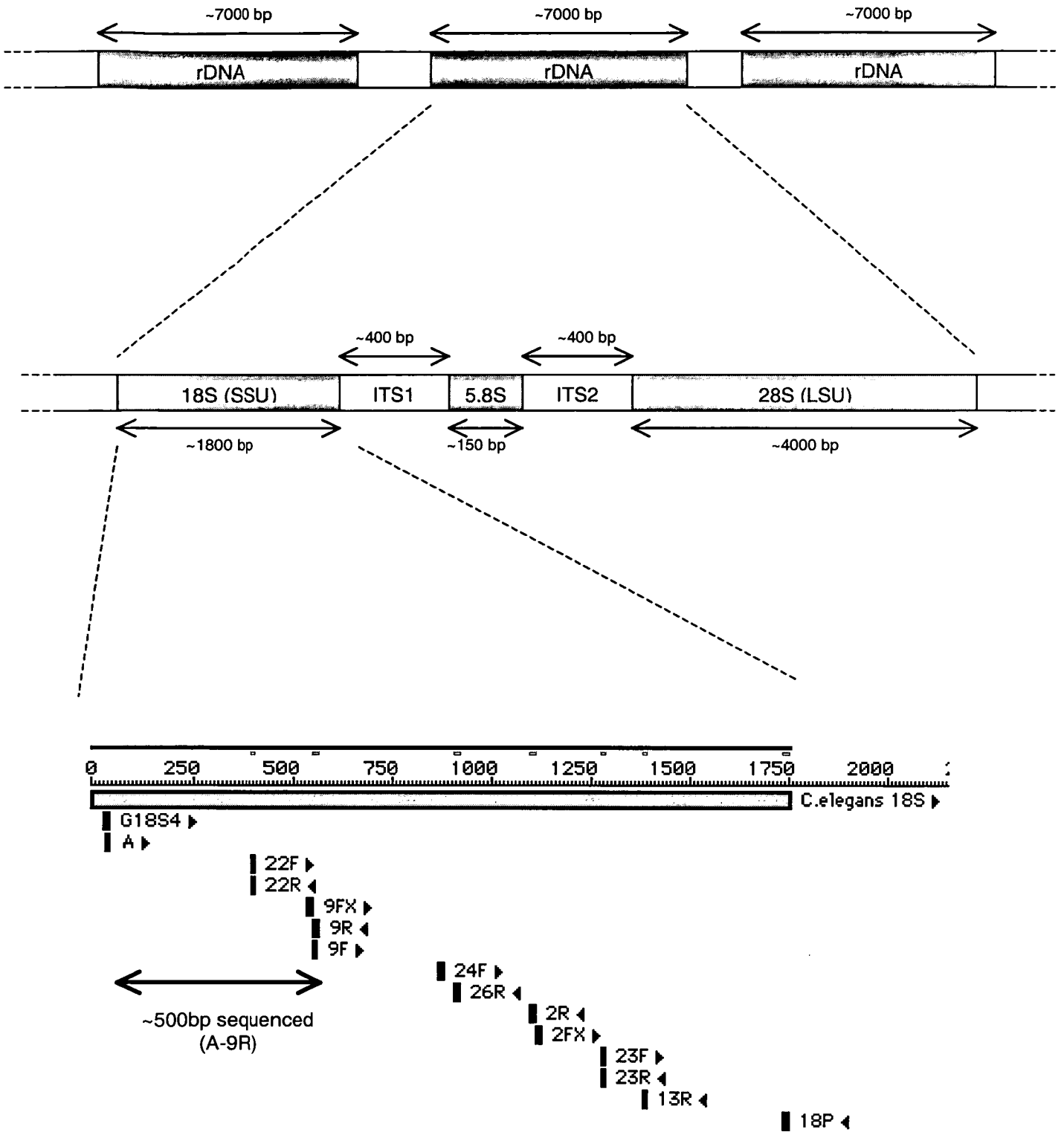



Figure 1.1 General structure of the eukaryotic nuclear ribosomal DNA cluster. Approximate lengths of each section are given in base pairs (bp). SSU = small subunit; ITS= internal transcribed spacer; LSU = large subunit. The lower part of the diagram shows the annealing sites of several primers within the SSU gene, and indicates the region chosen for sequencing in this project.

The internal transcribed spacer (ITS) region, which lies between the small and large subunit genes within the rDNA repeat, has also been proposed as a taxonomic marker; the ITS1 and ITS2 were tested for use as a sequence-based marker in this study, but were found to contain too high a level of variability to be useful: polymorphisms were found even between copies within the same individual, creating significant difficulties for sequencing, and the large degree of variation between taxa in both length and sequence would make alignment problematic for novel sequences.

The concept of a DNA-based taxonomic system which could be extended to biology as a whole has recently gained increasing support (Blaxter 2003; Blaxter and Floyd 2003; Hebert et al. 2003b; Tautz et al. 2003) but has also generated controversy (Lipscomb et al. 2003; Seberg et al. 2003); proponents suggest that DNA sequences should be obtained from all existing and future taxonomic samples and species descriptions, and that this sequence should serve as the basis for classification. Opponents point out problems including the costs of sequencing, difficulties in alignment of sequences, and uncertainty in distinguishing between orthologues and paralogues; some question whether a small segment of the genome can provide sufficient information for robust and meaningful taxon assignment. But regardless of the arguments one way or the other, such issues can only be resolved through testing the methodology in question in the real world and assessing its performance.

The mitochondrial cytochrome oxidase I (COI) gene has been proposed as a universal DNA barcode for animals (Hebert et al. 2003a), and was shown to be able to correctly place a sample of Lepidoptera to their correct, previously identified morphological species. However, as a mitochondrial gene its uniparental (female only) mode of transmission causes some evolutionary anomalies in certain groups, particularly insects with haplodiploid sex determination where females outnumber males, and thus mitochondrial genes can have far larger effective population sizes than nuclear equivalents (Navajas and Boursot 2003). The latter study also indicated significant problems of within-taxon variability, in that multiple distinct mitochondrial DNA lineages were found within a single taxon as defined by rRNA markers and by morphology.

For a universal DNA barcoding system it would perhaps be best not to rely on a single gene (Mallet and Willmott 2003), but to sequence two or more genomic regions - ideally with differing rates of evolution - from each specimen (or at least from each type specimen), providing a range of taxonomic resolutions. In any case, nematodes would appear to be an appropriate test group for pioneering a molecular biodiversity assessment method which might ultimately be applied to all life.

## 1.4 Use of molecular methods in biodiversity surveys

Recent years have seen a great expansion in the number of studies applying molecular approaches to ecological community analysis, particularly to groups of organisms which present difficulties to traditional taxonomy (reviewed by Theron and Cloete 2000; Nee 2003). For example, the study of bacteria,

fungi and other single-celled organisms was previously limited to those groups which could be grown in culture, as there was no other means to detect their presence. It has been estimated that more than 99% of prokaryotic species cannot be cultured (Woese 1996); therefore, knowledge of microbial diversity has been strongly biased towards a small subset. However, molecular techniques allow assessment of microbial diversity directly from environmental samples, requiring no culturing. A large number of such studies have now been published, and in virtually every case have resulted in the discovery of a vast and hitherto unknown diversity.

For example, Barns et al. (1999) used universal primers to amplify 16S rDNA (the prokaryotic SSU, equivalent to 18S in eukaryotes) from environmental samples including marine sediment, terrestrial soil, aerosol, hot spring and animal faeces. Numerous sequences clustered into a novel group with only one known cultivated member, *Acidobacterium capsulatum*. Analysis of these sequences showed that the group is as phylogenetically diverse and distinct as previously recognised bacterial divisions, and therefore constitute a previously unknown, major bacterial lineage. Many similar studies have also uncovered unexpected prokaryotic diversity in a range of habitats, including marine bacterioplankton (Rappe et al. 2000), Siberian tundra soil (Zhou et al. 1997) and the anoxic soil surrounding rice roots (Großkopf et al. 1998). Indeed, it is now widely agreed among the microbiological community that DNA sequences are an appropriate method for delimiting species-level groups in prokaryotes, given the lack of morphology or other means of delineation. It has been proposed that a sequence similarity level of >97% in 16S rDNA genes "is a reasonable level for grouping bacteria into species" (Hagstrom et al. 2002).

An immense diversity of microbial eukaryotes has also been revealed as a result of molecular analyses. Surveys of the 18S rDNA sequences of planktonic picoeukaryotes (Moreira and Lopez-Garcia 2002) have found a wide variety of lineages, many belonging to known photosynthetic classes, others to heterotrophic or mixotrophic classes; other sequence groups (termed 'phylotypes') were not clearly affiliated to any known organism. Additionally, a survey of deep-sea waters (250-3000m depth, i.e. below the photic zone) discovered a diverse community of picoeukaryotes – a significant discovery, as very little was previously known of deep-sea planktonic communities. A survey of the highly acidic and heavy metal-rich Rio Tinto, or 'River of Fire' in Spain (Amaral Zettler et al. 2002), discovered the surprising result that the 18S sequence diversity of eukaryotes is far greater than that of the prokaryotes previously known to exist in such extreme environments.

Even viruses, long considered intractable to ecosystem-level surveys, have proven accessible subjects for molecular analysis. Zhong et al. (2002) analysed viruses infecting marine cyanobacteria using sequences of the viral capsid assembly protein gene g20. Again, a high diversity was found: there were a total of 114 distinct g20 sequence types, falling into nine major phylogenetic groups, which were genetically divergent but more closely related to each other than to the outgroup, bacteriophage T4. Of these nine clusters, only three contained known cyanophage isolates; the identity of the other six remained unknown. The composition of the estuary and open ocean samples were also found to differ from each other.

While these studies demonstrate the utility of molecular approaches in studying entire communities, it has also been possible to examine specific groups within a community. For example, Salles et al. (2002) used primers specific to the bacterial genus *Burkholderia* to amplify 16S rDNA sequences from soil samples, then applied DGGE to examine the diversity of species within this genus. Using this method the authors were able to reveal differences in *Burkholderia* diversity between two grassland plots.

A range of novel methods have been applied to the study of microbial diversity in soil (reviewed by Torsvik and Øvreås 2002). The complete set of genomes in the soil community can be considered as one large genome – the 'metagenome' (Rondon et al. 2000) – which can itself be cloned in bacterial artificial chromosome (BAC) libraries, and its expression analysed using microarrays, providing a wealth of new information about genes and metabolic pathways.

For large land vertebrates, at least, we might expect that taxonomic diversity has been thoroughly catalogued. Traditionally, it is considered that there is one species of elephant in Africa, *Loxodonta africana*. Yet, suprisingly, studies of nuclear genes in African elephants have revealed at least two (Roca et al. 2001), and possibly three (Eggert et al. 2002) distinct taxa: forest and savannah elephants in Central Africa, with West African elephants possibly constituting a third lineage. Each group is approximately as divergent from one another as any is from the Asian elephant, *Elaphas maximus*. Having identified these taxa by DNA variations, it was shown that numerous morphological features and habitat distinctions also supported the establishment of new species, with implications for conservation policy. This demonstrates that even in groups where we might expect our taxononic inventories to be nearly complete, molecular analyses can still serve to reveal hitherto unnoticed diversity.

Within nematodes, molecular phylogenetic methods have also been used to help resolve species within taxonomically "difficult" groups (Adams et al. 1998; Beckenbach et al. 1999; De Ley et al. 1999). In this last study, it was shown that a pair of nematode strains within the genus *Acrobeloides*, which differed only in their body "handedness" (i.e. one strain formed a physical mirror-image of the other) and were otherwise morphologically identical, were nevertheless reproductively isolated from one another, and could also be distinguished by variations in their LSU rDNA sequence. Thus a pair of species (according to the BSC), which proved very difficult to distinguish by morphology, could be separated by DNA sequence identity.

Nematode diversity studies have also begun to be carried out using molecular methods, such as DGGE (Foucher and Wilson 2002). RFLPs and AFLPs have also been used as markers to distinguish species within particular nematode groups, such as the cyst (Szalanski et al. 1997) and root knot nematodes (Powers and Harris 1993; Semblat et al. 2000). Markmann (2000) analysed the diversity of 28S (LSU) rDNA sequences in lacustrine meiofauna, including nematodes. However, to date no large-scale survey of molecular sequence diversity has been applied specifically to nematodes. If, as Lambshead (1993) suggests, nematodes, like bacteria, are a group concealing a vast hidden diversity, DNA sequencing offers a means by which we might detect and describe this diversity.

## 1.5 Free-living nematode diversity: what is known?

One of the best known patterns in ecology is the relationship between latitude and diversity. For a very large proportion of organisms of all phyla, terrestrial and aquatic, species richness is maximal at the equator, and decreases towards the poles (Rosenzweig 1995). Certain data suggest that free-living nematodes may be an exception to this rule – a survey of North Atlantic marine nematodes found a significant increase in species richness moving northwards from the equator (Lambshead et al. 2000). It was believed that soil nematodes were also an exception to the general pattern (Procter 1990), as they appeared to show greater species richness at higher latitudes, with the maximum diversity and abundance in temperate regions, and were relatively unimportant in tropical soil faunas. However, more recent findings have called this conclusion into question (Boag and Yeates 1998; Ettema 1998). It is likely that differences in sampling intensity, rather than genuine patterns of diversity, were responsible for the perceived disparity between temperate and tropical soils, as recent surveys of tropical rainforest soils in Cameroon have revealed an immense diversity of morphospecies (Bloemers et al. 1997; Lawton et al. 1998). A summary of several studies is given in Table 1.1.

| Ecosystem | Latitude | Survey area (ha) | Total samples | No. morphospecies | Reference |
|---|---|---|---|---|---|
| | | | | | |
| Polar (Antarctica) | 180° | ? | 130 | 3 | (Freckman and Virginia 1997) |
| Subarctic (Sweden) | 59° | ? | 1 | 34 | (Ruess et al. 1998) |
| Agricultural (Tennessee) | 35° | ? | 30 | 100 | (Baird and Bernard 1984) |
| Hardwood forest (Indiana) | 40° | 118 | 54 | 175 | (Johnson et al. 1972) |
| Pasture (Denmark) | 55° | ? | 25 | 226 | (Overgaard Nielsen 1949) |
| Prarie (Kansas) | 38° | 259 | 61 | 228 | (Orr and Dickerson 1966) |
| Rainforest (Cameroon) | 3° | 24 | 24 | 431 | (Bloemers et al. 1997) |

Table 1.1 Soil nematode species diversity recorded in various ecosystems (from Ettema 1998).

This table also illustrates the considerable differences in sampling strategies employed by different surveys, making direct comparisons between different ecosystems problematic. It is well known that spatial scale has a significant effect on measured diversity. For example, a square metre of European grassland is more diverse in plant species than a square metre of Tropical rainforest, but for a square kilometre, the reverse is true (Groombridge 1992). Thus the high diversity associated with certain sites may simply reflect scale of sampling. More data on nematode abundance and diversity, with a more standardised methodology, are needed in order to draw conclusions about global-scale patterns.

A further difficulty which has become increasingly apparent is that the true number of undescribed species of nematode is far greater than was previously thought. Lambshead (1993) estimated the global diversity of marine nematodes by extrapolating from the number of new species found in each square kilometer, and arrived at a figure of 100,000,000 species. Clearly the reliability of such a figure is highly dependent on initial assumptions, but it is evident that the total of 4000 described species of marine nematode (Lambshead 2001) is a significant underrepresentation. A nematode survey in Mbalmayo Forest Reserve, Cameroon found a total of 374 nematode morphospecies, only 10% of which could be assigned to known species (Lawton et al. 1998). This same study surveyed additional taxa including birds, beetles, ants and termites; among these, the nematodes both showed the greatest proportion of unknown species, and required by far the greatest amount of scientist-time to sample, sort and identify.

## 1.6 Nematodes in ecosystem processes

Nematodes are known to play a variety of roles in soil ecosystem processes (Freckman 1982). They occupy a range of trophic ecologies, including bacteriovores, fungivores, plant browsers, plant parasites, animal parasites with free-living dispersal stages, and carnivores feeding on other soil animals including nematodes. However, our knowledge of the feeding behaviour of specific taxonomic groups is incomplete (Yeates et al. 1993).

The use of nematodes as 'bioindicators' of environmental disturbance was developed into a systematic protocol by Bongers (1990). Nematode families were placed on a five-point scale from "colonisers" (rapidly reproducing, quickly dominating unstable or ephemeral habitats) to "persisters" (slow life cycle and reproductive rate, found in habitats after long periods of stability), and used to calculate the Maturity Index (MI), the weighted mean of the individual coloniser-persister (c-p) values in a sample. A habitat recently disturbed by pollution or other environmental changes has a low MI value as its nematode fauna is dominated by colonizers, and the MI value increases during the process of ecological succession as more persisters appear.

Experiments linking nematode diversity to ecosystem function have been limited, but one such study (Ruess et al. 2001) examined the effect of soil manipulations on nematode trophic groups. At two contrasting arctic sites, soil was manipulated by NPK fertilization and addition of labile carbon (sucrose),

and bactericides and fungicides were used to manipulate bacterial and fungal biomass. Bactericide treatment was found to have no effect on the nematode fauna, but was also revealed to have very little effect on bacteria, due to the low persistence of the antibiotics in the soil. Conversely, the fungicide benomyl caused nematode species richness and maturity index values to decrease, but proved to be a strong toxin to nematodes as well as fungi, therefore this effect cannot be attributed solely to the loss of fungi. Both nutrient and energy enrichment were also found to cause decreases in species richness and the maturity index; however, although species numbers and abundances of individual species changed significantly, trophic structure was affected only slightly. That is, if one group of plant feeders decreased in abundance as a result of a certain treatment, another would increase, so that the overall proportion of plant feeders remained approximately constant. This suggests a degree of functional redundancy among nematode species.

A recent study (Ekschmitt et al. 2001) tested various aspects of soil nematode community structure and examined their correlation with a series of ecological parameters. Nematode abundance, biomass, respiration and species richness, as well as the MI value, were tested against microbial biomass and respiration, soil $NH_4$, $NO_3$ and organic nitrogen. Species richness was found to strongly correlate with the microbial parameters, while the MI value was correlated with nitrogen status. It was concluded that nematodes were potentially a useful indicator of soil function, but could only be considered more efficient than a direct measurement of soil parameters of interest if nematode identification could be accomplished more rapidly and simply, given the great expertise and time required for any morphological-based survey.

A series of studies have also examined the relationships between community structure and ecosystem processes by experimental manipulation of the former. Artificial microcosms of sterilised soil to which known species of bacteria, fungi and nematode were added (Mikola 1998; Mikola and Setälä 1998). Food webs were created with either one, two or three trophic levels, and studied over a five month period. Twenty species of bacteria and fungi formed the first trophic level, a bacteriovorous nematode (*Caenorhabditis elegans*) and a fungivorous nematode (*Aphelenchoides* sp.) the second, and a predatory nematode (*Prionchulus punctatus*) the third. It was found that microbivore biomass was regulated by the predator (biomasses of both microbivorous nematodes were significantly higher in the absence of *Prionchulus*); that microbial production was higher in the food webs with two and three trophic levels than when the microbes were growing alone; however, neither microbial biomass nor respiration were affected by the reduction in biomass at the second level due to the predator. Net mineralization of C and N was highest in the food chains with two trophic levels, at an intermediate level in the presence of predators, and lowest in the microbial-only communities. Thus, even in this relatively simple model ecosystem, a considerably complex system of interactions and regulations is taking place.

## 1.7 Aims of Project

The principal aims of this project were to devise and prove a molecular method for nematode diversity assessment, and to apply it to the analysis of the nematode fauna at a chosen study site (Fasset Hill, Sourhope, near Kelso) – thereby determining the total number of taxa present, and testing for correlations in the distribution of those taxa with various environmental parameters. This has involved developing and testing a series of laboratory methods allowing determination of DNA sequences from nematode samples, computational methods to robustly assign a set of sequences to molecular taxa, and statistical methods to analyse patterns in taxon diversity. Additionally, work done by Dr Eyualem Abebe, who carried out a morphological survey in parallel with the molecular survey, have allowed the findings of these two approaches to be directly compared.

This thesis will describe these methods and the results of the molecular survey, will discuss the issues involved in defining taxa by DNA sequences, and will interpret what conclusions may be drawn from these findings and how they fit into the larger issues of biodiversity research.

# 2. Methods, part 1 – field and laboratory

## 2.1 Study Site

The site chosen for this project was Sourhope farm on Fasset Hill, near Kelso in the Scottish Borders (grid reference NT 620 384, altitude ~260m). The site is a grassland field of soil type U4 in the UK soils classification, dominated by *Agrostis* and *Festuca* grass species; it is part of the UK Environmental Change Network and is also the main study site for NERC's Soil Biodiversity Programme (see http://mwnta.nmw.ac.uk/soilbio/index.html). An extensive set of environmental data relating to the site has been collected by site management staff and other groups within the Programme, including vegetation biomass and species abundances, soil moisture, soil pH, and site topography.

The dimensions of the experimental area are approximately 80 m by 115 m. As shown in Figure 2.1, the site is divided into 30 plots of 12 m by 20 m, in which five different treatments are replicated: control, liming (1.2 kg $m^{-2}$ added once a year), nitrogen addition (12 g $m^{-2}$ added once a year), nitrogen addition together with liming (both previous treatments), biocide (Dursban – 36 ml in 10 litres per plot added once a year). A second set of control plots was present but not used in this project. Each plot is also divided into subplots (see Figure 2.2). Our project was allocated four subplots from which to sample (S, T, U and V) from which four replicate soil cores per plot were taken in each sampling event.

After preliminary samplings in 2000 and 2001 to test and optimise various laboratory methods, the two main surveys were carried out in June (control plots only) and October (all plots) of 2001. Soil was sampled using a 4 cm diameter soil corer, to a depth of approximately 10 cm; total soil depth across the site is ~10 to 20 cm. June soil cores were divided into upper (organic) and lower (mineral) horizons of around 5 cm each, based in the visible change in soil colour. October cores were left undivided to reduce the number of samples needing to be processed. Soil samples were stored in polythene bags at 4°C until used.

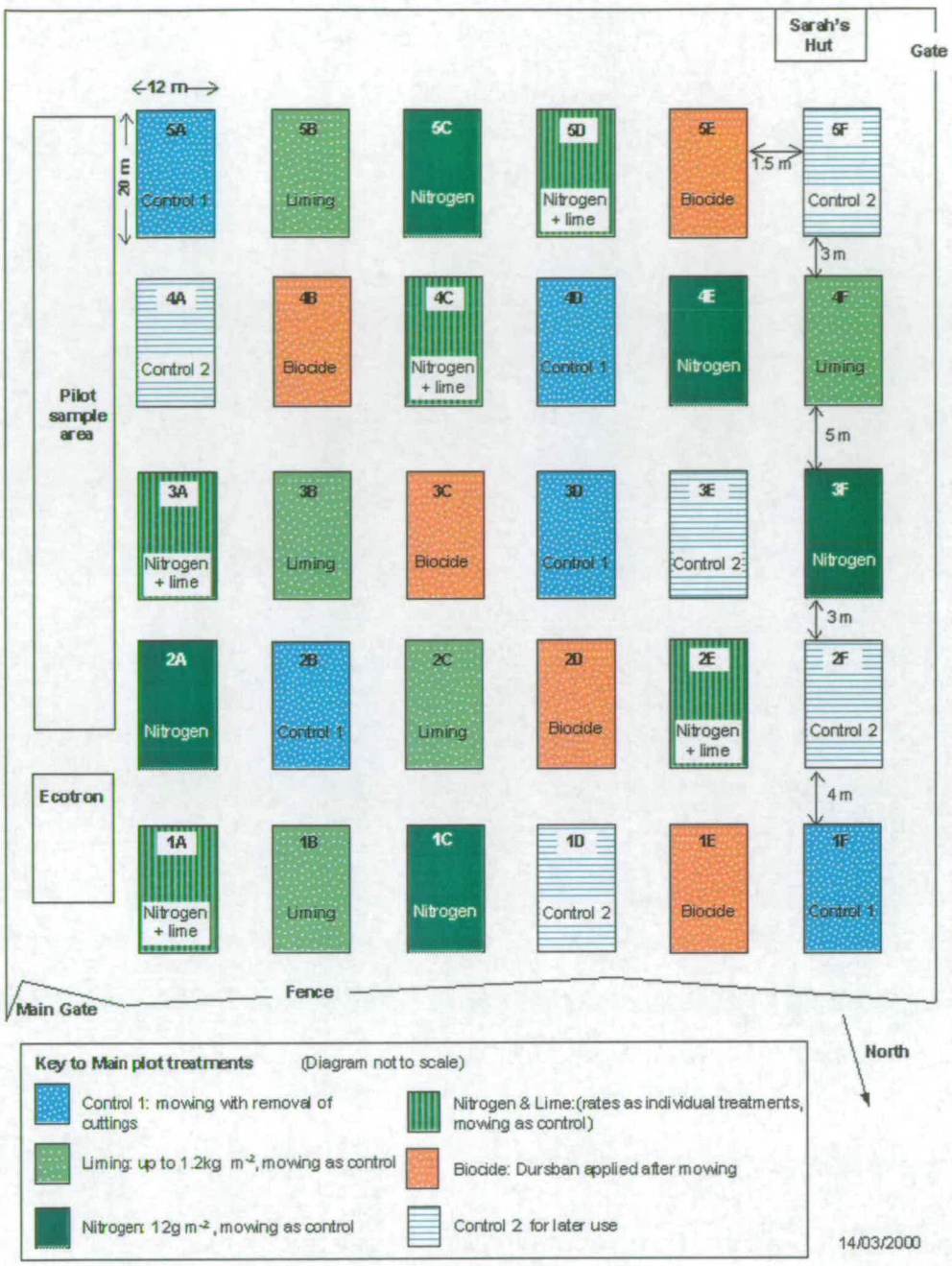



Figure 2.1: Sourhope field layout, showing main plots and treatment allocations (image from NERC Soil Biodiversity Programme website - http://mwnta.nmw.ac.uk/soilbio/Sourhope.htm)

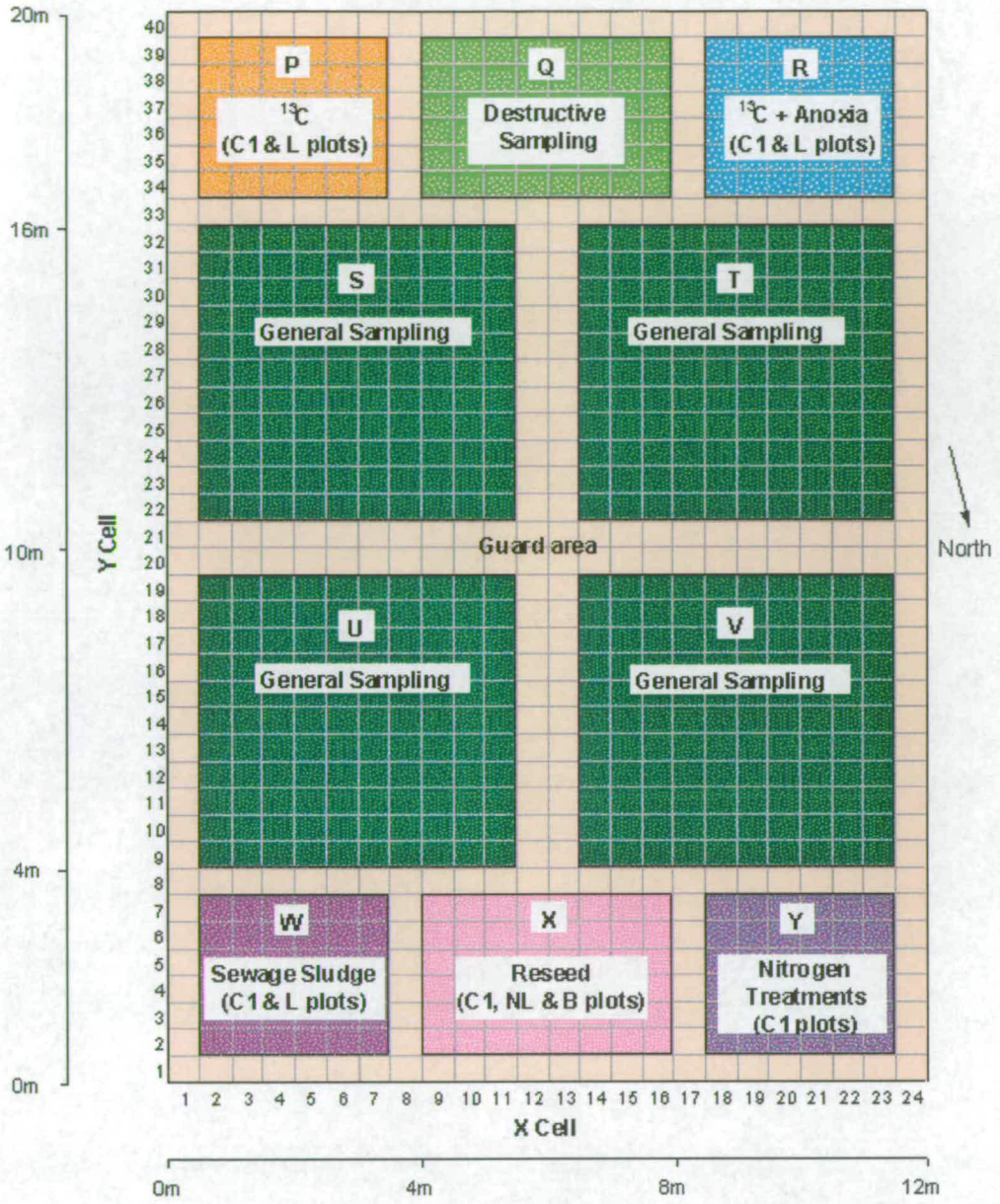



Figure 2.2: subplot design within each plot (image from NERC Soil Biodiversity Programme website - http://mwnta.nmw.ac.uk/soilbio/Sourhope.htm)

## 2.2 Nematode extraction

Each soil sample was transferred to a 500 ml centrifuge bottle. Approximately 200 ml of sterilised tap water was added to bring each sample to equal weights. One teaspoon of kaolin (hydrated aluminium silicate, supplied by Sigma Aldrich) powder was added, aiding in the settling of detritus and sediment. Samples were mixed by shaking and then centrifuged at 400 x g for 5 minutes; the supernatant (containing soluble substances and light plant material) was discarded. Approximately 200 ml of Ludox (colloidal silica, supplied by Sigma Aldrich) with a specific gravity of 1.15 was added, and samples were thoroughly mixed by shaking. Samples were spun again at 400 x g for 5 minutes, settling detritus and sediment to the bottom, but leaving all material less dense than the Ludox (including nematodes and other meiofauna) floating in the supernatant.

The supernatant was poured onto a 38 μm mesh sieve, and washed with a water spray; Ludox and small particulate material was washed through, but nematodes were trapped on the sieve. Nematodes were then washed off the sieve and collected in a plastic bottle in a clean suspension of tap water. Each sample was then split into two bottles: one was bulk-fixed in hot formalin so that nematodes could be picked to slides for morphological identification; from the other, nematodes were picked individually to be processed for PCR. In an effort to ensure sampling of individual specimens was as random as possible, nematodes were picked from a counting dish with a grid on its base. Under a binocular microscope, each square on the grid was examined in turn and the nematode lying at the centre of each square, or closest to the centre, was selected and transferred using a platinum wire pick to a single 0.2 ml tube (usually as part of a 96-well plate) for DNA extraction.

## 2.3 DNA extraction

Nematodes were digested to release genomic DNA following the method of Stanton et al. (1998), slightly modified to incorporate a shorter digestion time which was found to give optimal results from these specimens. Each nematode was placed in a tube containing 20 μl of 0.25 M sodium hydroxide, which was then incubated at 25°C for between 3 and 16 hours to allow digestion of the nematode (any digestion time within this range was found to give approximately equal PCR success rates). The samples were then heated for 3 minutes at 99°C. 4 μl of 1 M HCl, 10 μl of 0.5 M Tris-HCl (pH 8.0), and 5 μl of Triton X-100 were added, and the samples were heated again for 3 minutes at 99°C (HCl neutralises the NaOH; Tris is a pH buffer; and Triton X-100 is a detergent which causes proteins to dissociate; the two heating stages denature proteins). This leaves each nematode digested in a final volume of 39 μl, with its genomic DNA stably buffered and available for PCR. Lysates were stored at -80°C and archived for future use.

## 2.4 Polymerase Chain Reaction (PCR)

1 µl of each nematode lysate was added to a 20 µl PCR reaction in a microtitre plate comprising Expand LT buffer 3 at 1x concentration; 2.25 mM $MgCl_2$; 0.2 mM each nucleotide; 0.7 units of Expand LT polymerase (both enzyme and buffer supplied by Roche Biochemicals); and 8 pmol of each primer. The primers used were SSU18A (AAAGATTAAGCCATGCATG) and SSU26R (CATTCTTGGCAAATGCTTTCG) (Blaxter et al. 1998), giving a ~1000 base pair PCR product (see Figure 1.1 for positions of primer binding sites within the SSU molecule). The reaction conditions were: (i) 94°C for 5 minutes; (ii) 28 cycles of {94°C for 1 minute; 55°C for 1 minute 30 seconds; 68°C for 2 minutes}; (iii) 68°C for 10 minutes. 5 µl of each product were run on a 1% agarose-TBE gel with ethidium bromide to determine if bands of the correct size were present.

Both the digestion and PCR protocols were optimised by testing on nematodes from monocultures isolated from Sourhope soil samples. Since only 1 µl of each nematode digest was required to generate SSU PCR product, a single nematode provides sufficient DNA for up to 39 PCRs. It was also found that 0.5 µl of lysate was often sufficient for PCR, albeit with a slightly lower success rate – perhaps related to variations in the size of the nematodes. For this reason 1 µl lysate per 20 µl PCR was used as the standard protocol for all survey nematodes.

Successful PCR products were prepared for sequencing by treatment with Exonuclease I (ExoI), which breaks down single-stranded DNA to dNTPs, thus removing unincorporated primers, and Shrimp Alkaline Phosphatase (SAP), which reduces nucleotide triphosphates to monophosphates, so that excess dNTPs do not affect subsequent sequencing (Werle et al. 1994). 1 µl of SAP and 1.5 µl of ExoI (both supplied by Amersham Biosciences) were added to 15 µl PCR product; reactions were heated at 37°C for 40 minutes and 94°C for 15 minutes.

## 2.5 Sequencing

SAP/ExoI-cleaned PCR products were sequenced using the DYEnamic ET terminator system (Amersham Biosciences). 2 µl of PCR product were added to 4 µl DYEnamic ET termiators, 1 µl primer at 5 pmol/µl concentration, and 3 µl of ultrapure water to make a total volume of 10 µl. Reactions were heated for 25 cycles of {95°C for 20 seconds, 50°C for 15 seconds, 60°C for 60 seconds}. The internal primer SSU9R (AGCTGGAATTACCGCGGCTG) (Blaxter et al. 1998), was used to generate approximately 500 bases of sequence at the 5' end of the molecule. SSU18A, the 5' primer used for PCR, was also used for some initial samples, but was found to give a lower success rate. Completed reactions were then cleaned using Amersham filtration columns, and run on an Applied Biosystems 377 automated sequencer.

It was important that soil samples were processed as quickly as possible after being brought back to the lab, to avoid the possibility that certain nematodes might die if left in soil at 4°C for too long. Therefore, three workers collaborated in picking the nematodes from the June and October samples: myself, Dr Eyualem Abebe (Postdoc), and Mark Welsh (a Reseach Assistant in the lab). For the June samples, Dr Abebe also assited in carrying out the PCR and sequencing reactions, so that all of the samples were finished with in time for the October sampling (for the October samples, all PCR and sequencing was carried out by me, though Dr Eyualem again assisted with picking nematodes). Appendix 1 shows a list of all sequences generated in this project, and indicates which worker was responsible for picking the nematode, carrying out the PCR and sequencing the PCR product.

# 3. Methods, part 2 – informatics

## 3.1 Processing of sequences

Sequence traces were analysed using the automated base-calling algorithm phred (Ewing and Green 1998; Ewing et al. 1998), which examines DNA sequencer chromatogram files, determines the most probable base at each position, and assigns a quality value (QV) to each; this is a measure of the probability that the base is called incorrectly, defined by the formula:

$$QV = -10 * \log_{10}(P_e)$$

where $P_e$ is the probability that the base call is an error.

Low quality regions of sequence (with a QV below 15) were trimmed, typically resulting in sequences of between 450 and 500 bp (see Chapter 4 for details); any file found to contain less than 400 bp of high-quality sequence after trimming was considered a 'fail' and excluded from subsequent analyses. Long sequences were also trimmed to a constant length defined by conserved sites close to the SSU18A and SSU9R primer sites, using the PERL script **primer_trim.pl** (included in Appendix 3). This script reads a set of sequences, searches for a specified pair of nucleotide strings at the 3' and 5' ends, and deletes any sequence outside these sites; it also flags any sequences for which a match was not found, so that these can be examined separately. A non-match may result either because the sequence is too short and ends before reaching the site in question, or because the sequence contains a variation within the site being searched for, in which case the excess sequence must be deleted manually at the appropriate position, determined by alignment with other sequences. The mean sequence length after trimming was 475 nucleotides.

## 3.2 Assignment of sequences to MOTU

The PERL script **define_MOTU.pl** (included in Appendix 3) was written for the purpose of clustering sequences into MOTU based on numbers of identical bases. This script is a modification of the sequence-clustering algorithm **CLOBB**, orignally written by John Parkinson (Parkinson et al. 2002). The new script takes a set of input sequences (as FASTA-format text files) and carries out a series of BLAST searches (Altschul et al. 1990; Altschul et al. 1997), so that every sequence is eventually searched against every other sequence, in a randomised order generated each time the script is run. From each search the top HSP (high-scoring segment pair) is extracted, along with the number of identical bases and the overall match length between the two sequences (i.e. the sequence currently being tested and its closest match). If the sequences are identical, or if the number of non-identical bases is equal to or below a user-specified threshold level (for example, 2 differing bases out of 475, or 99.58% identical), the sequences are assigned to the same MOTU; if the level of difference is above the threshold, they are assigned to different MOTU.

Ambiguous characters such as gaps and unresolved base calls were ignored in this analysis (i.e. the number of gaps and Ns in a pair of sequences was subtracted from the match length) – this was done to allow the inclusion of potentially noisy data, so that new MOTU were not defined solely on the basis of sequencing errors or regions with problematic alignments. Each MOTU is given a unique name indicating the level of similarity used to define it, and a four digit identifying number, for example '2bp_MOTU0001'.

A single consensus sequence was then derived for each MOTU using the DNA fragment assembly program phrap (see http://www.phrap.org/; Ewing and Green 1998; Ewing et al. 1998). Consensus sequences were compared using BLAST to a custom database of existing SSU sequences from nematodes and other organisms, to tenatively assign conventional taxonomic names to the nematodes from which the MOTU sequences had been generated, and also to screen for any sequences derived from organisms other than nematodes. Inevitably, since nematodes were picked directly from soil extracts containing a multitude of soil biota, the original nematode digests from which PCR products were generated undoubtedly contained DNA from organisms other than nematodes, some of which could potentially act as a PCR template for the primers used. For example, a number of sequences showed a high similarity to the SSU of the soil fungus *Mortierella chlamydospora*, suggesting that in these instances an SSU gene from a fungus, presumably carried with the nematode when it was picked from soil suspension, had been PCR-amplified instead of the nematode gene. All sequences identified as being of probable non-nematode origin were excluded from subsequent analyses.

## 3.3 Phylogenetic Analysis

MOTU consensus sequences were aligned using ClustalW (Thompson et al. 1994; Thompson et al. 1997), together with a set of sequences from known nematode taxa. Phylogenetic analysis was carried out using PAUP* 4.0b10 (Swofford et al. 1996; Swofford 1999). To visualise the similarities of MOTU sequences to each other and to known taxa, trees were constructed using the neighbour joining algorithm with absolute number of character differences as the distance measure (i.e. no correction for multiple substitutions at the same site, or for unequal transition/transversion rates).

## 3.4 Relational database design

A series of tables was generated collating information about sequences, MOTU, and sites of origin. From these a relational database was constructed using PostgreSQL, as represented in Figure 3.1. This database links every SSU sequence from an individual nematode (each identified by a unique name) to a specific MOTU to which that sequence belongs, and also to information about where and when the nematode was sampled; additionally, the database links each MOTU to taxonomic and biological information. SQL queries allow searching on any of these properties. Assignment of specimens to higher

taxonomic groups was done following the sytem of De Ley and Blaxter (2001); maturity index (MI) values for each nematode family were taken from Bongers (1990), and trophic group information from Yeates et al. (1993).

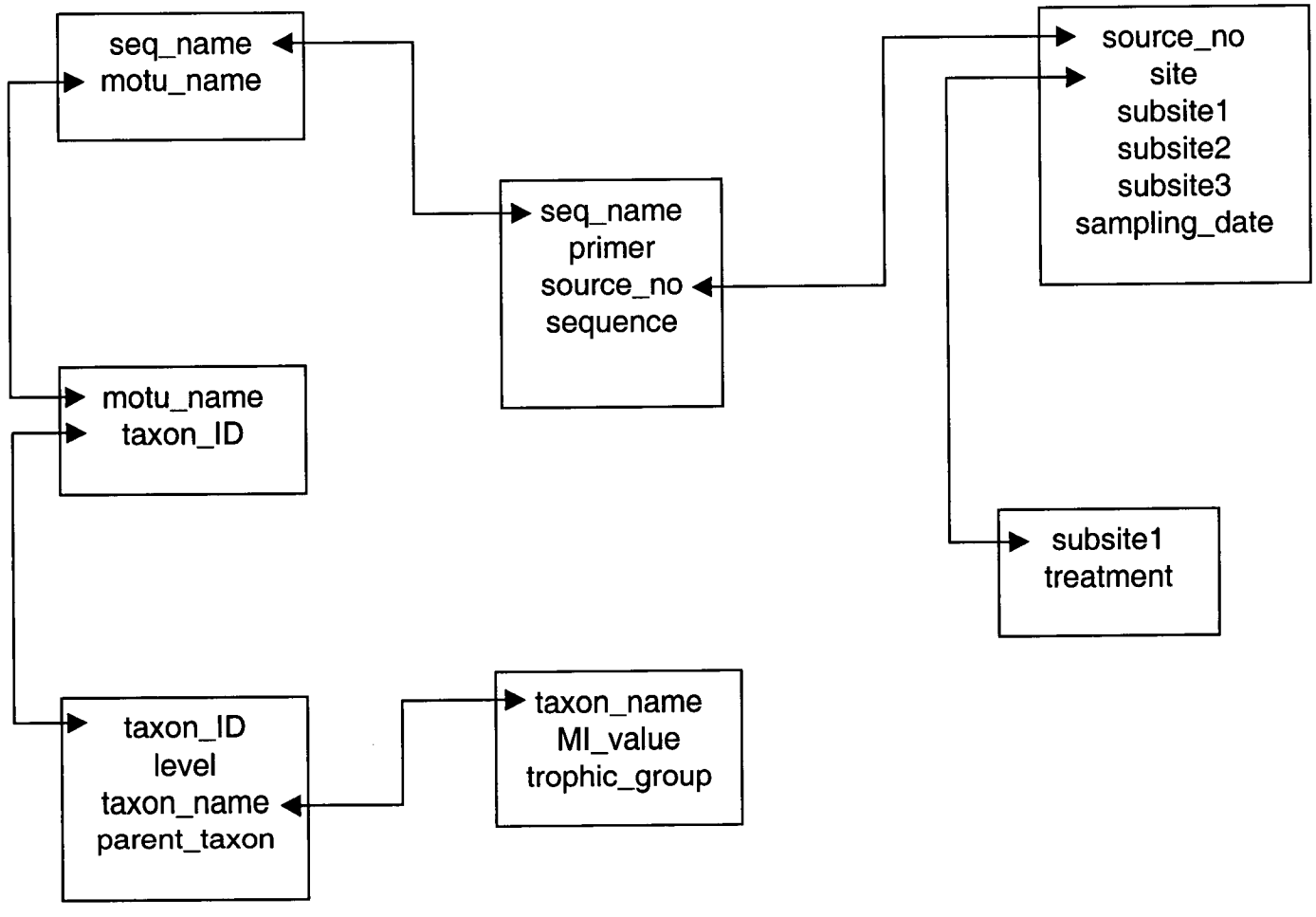



Figure 3.1: see next page for full legend.

Figure 3.1: design of SQL database. Each box represents a table of data, which are linked by matches (keys) in the fields joined by double-headed arrows.

seq_name = a unique name assigned to each sequence from an individual nematode, e.g. 12345ED

motu_name = name for each MOTU, e.g. 2bp_MOTU0001.

primer = which primer was used in sequencing (18A or 9R)

sequence = the nucleotide sequence itself, after trimming for quality

source_no = a unique number assigned to each sequence indicating where and when it was sampled

site = the location from which a nematode was sampled – this is "Sourhope" for all data in this project, but eventually could include other sites.

subsite1 = 1st subdivision of the sampling site – for Sourhope data this is the plot name – 1A, 1B, etc.

subsite2 = 2nd subdivision - for Sourhope data this is the subplot (S, T, U or V) where known.

subsite3 = 3rd subdivision - for Sourhope data this is the upper or lower horizon, for samples which were split.

sampling_date = date of sampling.

treatment = experimental treatment applied to each plot.

taxon_ID = a unique number assigned to each traditional named morphotaxon

level = the systematic level of the taxon (species, genus, etc.)

taxon_name = the name of the morphotaxon.

parent_taxon = the taxon immediately above the current taxon (i.e. the family to which each genus belongs, the order to which each family belongs, etc.)

MI_value = maturity index (coloniser-persister scale) value for each taxon, where applicable (Bongers 1990).

trophic_group = the trophic ecology of each taxon (bacteriovore, fungivore, etc.), where known (Yeates et al. 1993).

# 4. Results - Overview

## 4.1 Preliminary survey

A small preliminary survey was carried out on soil samples collected in July 2000, as a test of the methods and a demonstration of the viability of the MOTU concept. 74 sequences were generated from nematodes randomly sampled from across the site, from which 19 MOTU were defined. A subset of 18 individuals sampled from a single subplot (4DU) yielded 8 MOTU, including 4 which were unique to this smaller sample. Additionally, a series of 166 cultured nematode strains were isolated from Sourhope soil collected in 1999 and 2000 by Dr Artemis Papert; all of these strains were both morphologically identified and sequenced (the standard ~500 bp SSU region) for assignment to MOTU. These 166 sequences clustered into only 5 MOTU. For full details of results see Floyd et al. (2002); this paper is included as Appendix 4.

## 4.2 Success Rate of PCR

The total number of nematodes picked and PCR-amplified in the June and October 2001 samplings was 3264. Of these, the number which were PCR positive (as judged by the presence of a visible band of the correct size on an ethidium gel) was 2690. The digestion/PCR step, therefore, had a success rate of approximately 82.4%.

In the June sampling, all of the control 1 plots (but no treatment plots) were sampled, and soil cores were divided into upper and lower horizons. 96 nematodes from each sample were picked for PCR, from a total of 10 samples (5 control plots x 2 horizons). The PCR success rate was found to vary between samples, as shown in Table 4.2.1 below. The minimum number of positives was 43, the maximum was 82, with an overall mean of 63.9 and a standard deviation of 12.2.

| Plot | Upper | Lower |
|---|---|---|
| | | |
| 1A | 72 | 61 |
| 2B | 66 | 65 |
| 3D | 79 | 48 |
| 4D | 64 | 82 |
| 5A | 59 | 43 |
| Mean | 68 | 59.8 |
| Standard deviation | 7.714 | 15.353 |

Table 4.2.1 Number of PCR positives from June 2001 sampling

| Plot | No. PCR +ve |
|---|---|
| | |
| 1A | 81 |
| 1B | 79 |
| 1C | 76 |
| 1E | 82 |
| 1F | 90 |
| 2A | 87 |
| 2B | 83 |
| 2C | 87 |
| 2D | 83 |
| 2E | 86 |
| 3A | 93 |
| 3B | 79 |
| 3C | 73 |
| 3D | 80 |
| 3F | 86 |
| 4B | 78 |
| 4C | 87 |
| 4D | 63 |
| 4E | 82 |
| 4F | 85 |
| 5A | - |
| 5B | 90 |
| 5C | 66 |
| 5D | 73 |
| 5E | 88 |
| Mean | 81.542 |
| Standard deviation | 7.378 |

Table 4.2.2 Number of PCR positives from October 2001 sampling

In the second main sampling, in October 2001, all plots (except Control 2) were sampled, but plots were not split into upper and lower horizons. Thus 25 soil cores were taken, from each of which 96 nematodes were picked for PCR (unfortunately the sample for 5A was lost, so no data are available from this plot). Again some variation was found in the success rate of PCR (see Table 4.2.2).

From these samples, the minimum number of positives was 63, the maximum was 93, and the mean was 81.5 ± 7.4 – a larger mean success rate and a smaller standard deviation than seen in the June samples. Since the October samples were processed later, it seems likely that the reason for the variation in success rate was technical rather than biological, and the improved results in later samplings can be explained by improvements in technique.

## 4.3 Success Rate of Sequencing

When these PCR products were sequenced, considerable variation was also found in the length of sequence produced. From the 2690 June and October PCRs which were sequenced, the number which yielded sequence data was 2303 (the remainder failed and produced no useable output). Combining these with the 85 sequences generated in the preliminary survey, Figure 4.3.1 shows the size distribution across all 2388 sequences produced from all samples, after trimming of low-quality sequence by phred (see Chapter 3 for details). The maximum length was 532 bases, the minimum was 150 (the cutoff below which phred, as it is configured on our system, considers a sequence a fail and produces no output); the mean was 458 bases, and the mode was 496. The maximum predicted length of sequence - based on the distance from the 3' end of the SSU18A primer site to the 5' end of the 9R primer site in the known SSU sequence of *Helicotylenchus dihystera* - is 535 bases.

Sequences which are too short cannot be used for placing individuals to OTU as they contain too little information. It was therefore necessary to define a cutoff length for 'useful' sequences, below which sequences are discarded. A minimum length of 400 bases was chosen as the cutoff point, as it was considered that these should contain sufficient information for taxon assignment while allowing the majority of sequences to be included; all sequences below this length (349 out of the 2388, around 15%) were excluded from subsequenct analysis, leaving 2061 sequences. After BLAST-searching against the SSU database, 22 of these were judged to be of non-nematode origin due to showing a top BLAST hit to an organism other than a nematode. After these were excluded, a final total of 2039 useable nematode sequences remained (74 from the preliminary survey, 1965 from the June and October samples); these are listed in Appendix 1. The mean sequence length after the exclusion of the short sequences was 485 ± 23.3. The total number of bases determined within these included sequences was 988002.

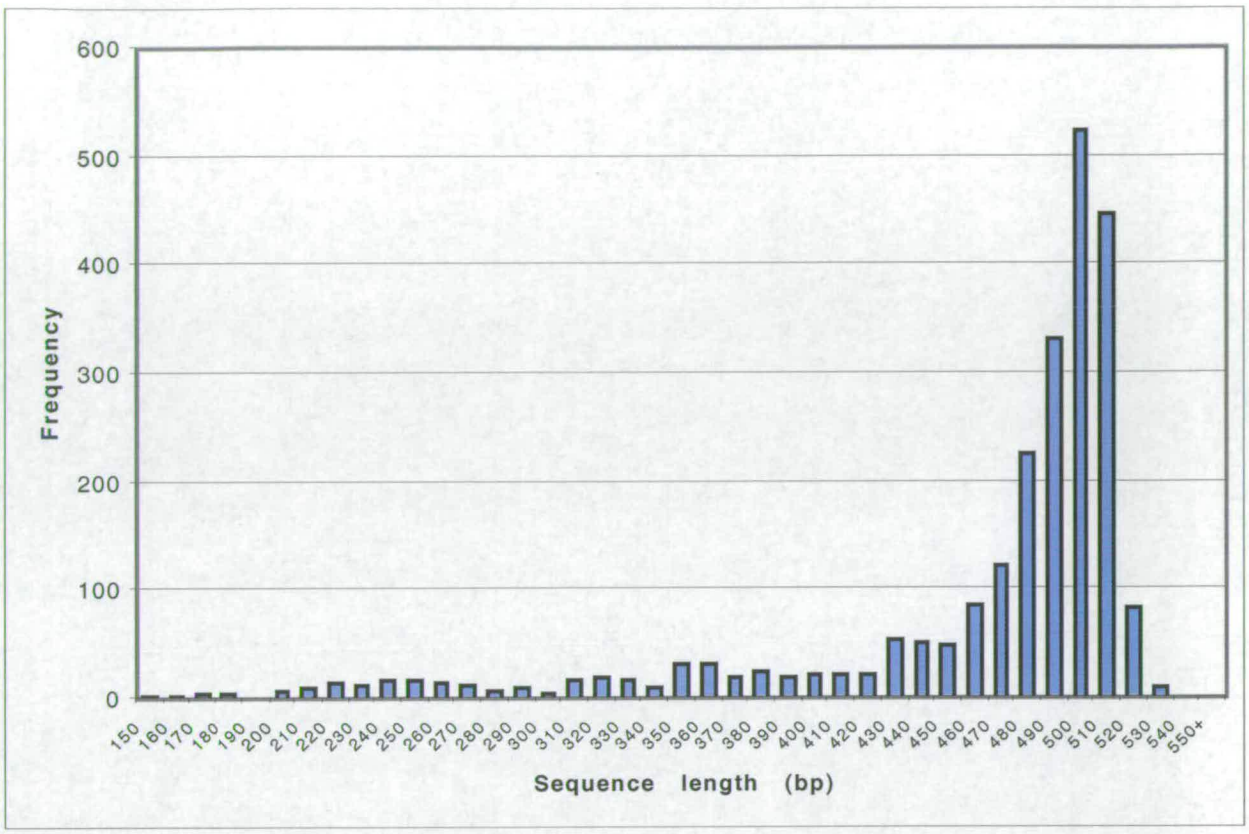



Figure 4.3.1 Histogram showing frequency distribution of lengths of all 2388 Sourhope survey sequences.

## 4.4 Processing of Sequences into MOTU

The PERL script **define_MOTU.pl** was used to cluster sequences into molecular operational taxonomic units (see Chapter 3). The 2039 sequences were found to cluster into 140 MOTU within which sequences differed by 2 bases or less (labelled '2bp_MOTU0001 - 2bp_MOTU0140'). The most abundant MOTU, 2bp_MOTU0002 (similar to the SSU of *Helicotylenchus dihystera*), contained 835 sequences, around 41% of the total; the majority of MOTU (88 out of 140) were represented by only a single sequence. The complete list of MOTU with their abundances is included in Appendix 2. However, when this process was re-run on the same set of sequences multiple times with sequence addition in a random order for each run, variations in MOTU assignment were found (see Chapter 5 for details).

## 4.5 Accumulation of MOTU

The abundances of MOTU represented by these 2039 sequences follow a common pattern seen in sampling many ecological communities: a small number of dominant taxa are represented by many individuals, while a large number of rare taxa are represented by a few individuals or only one (Rosenzweig 1995). Figure 4.5.1 shows a graph of $\log_e$(abundance) of each MOTU plotted against rank, while Figure 4.5.2 shows a frequency distribution graph of the same data.

If individuals are sampled at random from such a community, we would expect the rate of discovery of new taxa to decrease gradually with increasing numbers of individuals sampled, finally reaching an asymptote determined by the maximum number of taxa in the community, after which no more taxa are found however many individuals are sampled. The plot of the cumulative number of taxa recorded against a measure of sampling effort (such as number of individuals sampled) is termed the taxon accumulation curve (Southwood and Henderson 2000). Figure 4.5.3 shows the accumulation curve for the 140 MOTU defined from all 2039 Sourhope sequences, derived from a PERL script which parsed the output of **define_MOTU.pl** (a list of sequence names together with their MOTU assignments) one sequence at a time and recorded each time when a new taxon was assigned. This graph does not appear to be approaching an asymptote, and indeed is approximately linear for most of the last thousand individuals. This suggests that even by sampling over 2000 individuals we have not observed every taxon at Sourhope, and if sampling were continued it is likely that further taxa would be found.

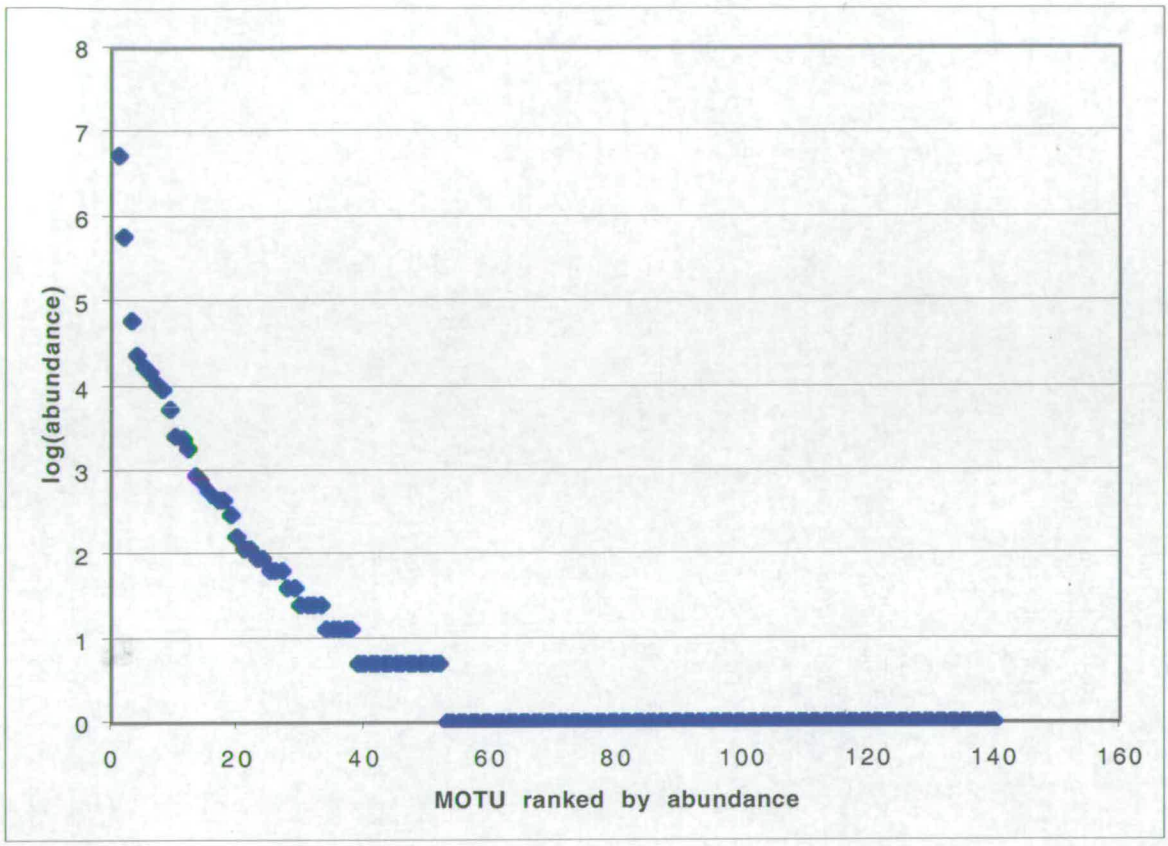



Figure 4.5.1 $\log_e$(abundance) of each MOTU plotted against rank by abundance.

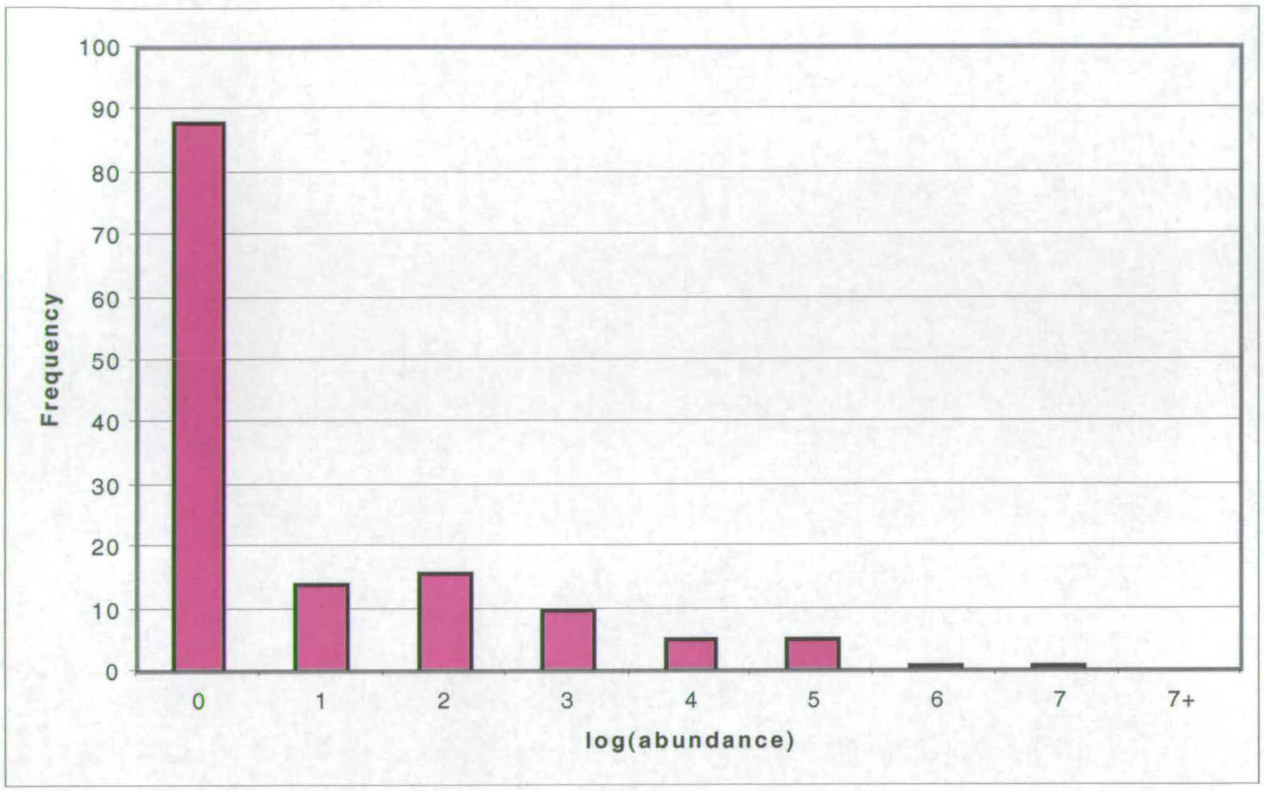



Figure 4.5.2 Distribution of $\log_e$(abundance) classes for all MOTU.

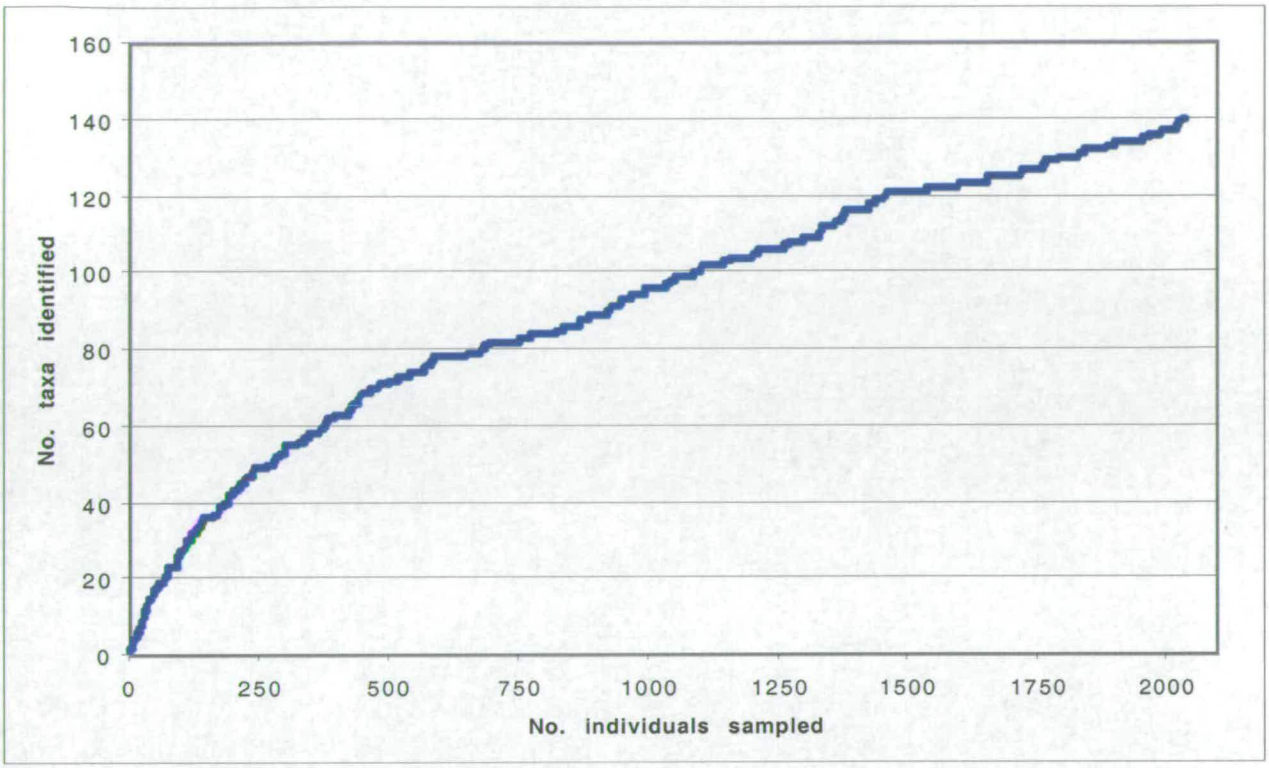



Figure 4.5.3 Taxon accumulation curve, plotting MOTU recorded by number of individual sequences sampled.

Even if an accumulation curve is not observed to reach an asymptote, the maximum number of taxa, $T_{max}$, can be estimated from parameters of this curve by various methods (Southwood and Henderson 2000). For example, a simple estimator was derived by Chao (1984), based on the total observed number of taxa ($T_{obs}$), and the number of taxa represented by one (a) and two (b) individuals:

$$T_{max} = T_{obs} + (a^2/2b)$$

For the current set of MOTU, $T_{obs}$=140, a=88, and b=14. Therefore the true maximum number of taxa at the Sourhope site is estimated to be approximately 417. If true, this suggests that the total of 140 taxa observed by sampling 2039 individuals represents only about a third of the true total at the site.

## 4.6 Distinctness of MOTU From Known Sequences

A single consensus sequence was generated for each MOTU (as described in Chapter 3), and the PERL script **base_diff.pl** was written to determine, for each MOTU consensus sequence, the number of bases (excluding gaps and Ns) by which it differed from the closest known nematode sequence by BLAST analysis (known sequences, including accession numbers and/or sources, are included in Appendix 1). The results are shown in Table 4.6.1 and and Figure 4.6.1.

| Base diffs. | Frequency |
|---|---|
| | |
| 0 | 8 |
| 1-5 | 35 |
| 6-10 | 47 |
| 11-15 | 23 |
| 16-20 | 12 |
| 21-25 | 2 |
| 26-30 | 8 |
| 31-35 | 3 |
| 36-40 | 0 |
| 41-45 | 1 |
| 46-50 | 1 |
| 50+ | 0 |
| | |
| Total | 140 |

Table 4.6.1 Numbers of bases by which each MOTU consensus differs from the nearest known nematode sequence (by BLAST analysis)

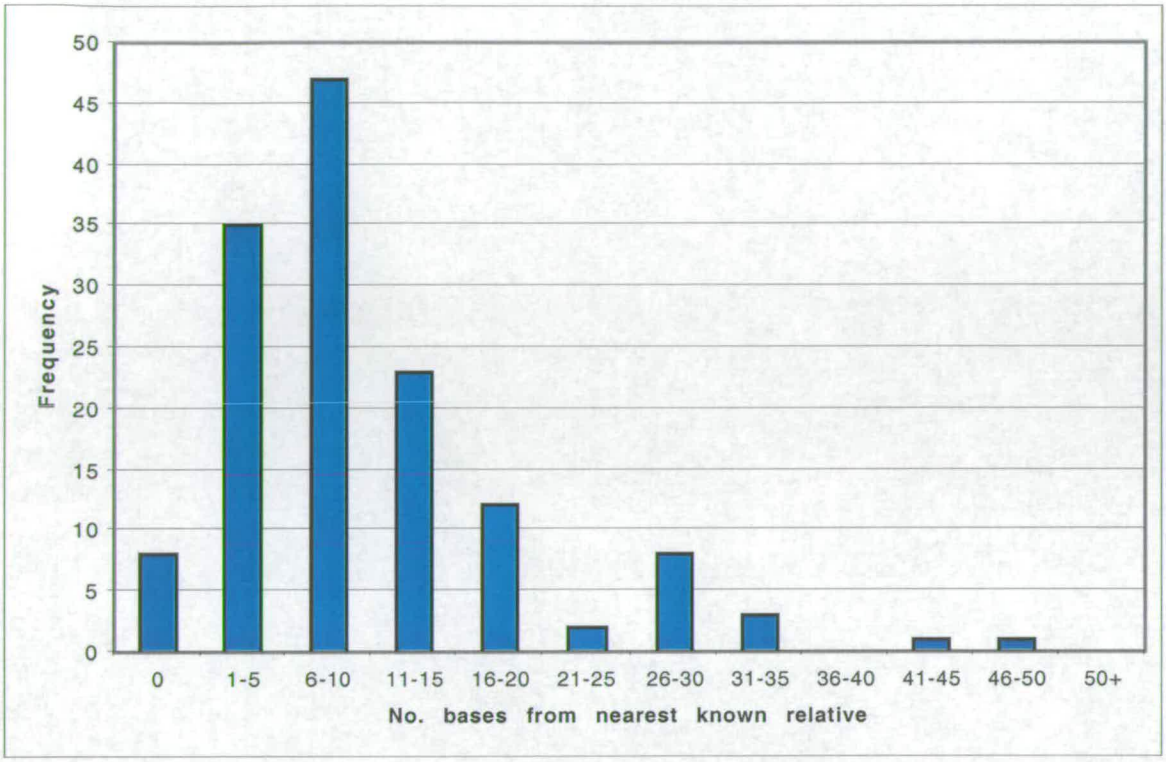



Figure 4.6.1 Histogram of numbers of bases by which each MOTU consensus differs from the nearest known nematode sequence (by BLAST analysis).

8 MOTU are identical to a known sequence over the region compared (discounting gaps and Ns); the majority of MOTU are between 1 and 10 bases from the nearest known sequence, with a progressively smaller number showing greater distance. 2bp_MOTU0037 is identical to both *Acrobeloides sp.* and *Cephalobus oryzae* (two closely related taxa which are not distinguishable by this method, but which are also well known as problematic taxa for morphological distinction). Most sequences recorded in this survey are novel, in the sense that they are not identical to any existing SSU sequence, though nearly all show sufficient similarity to some existing sequence to allow tentative taxonomic identification. The maximum numbers of differences are 41, shown by a MOTU closest to *Prismatolaimus intermedius*, and 48, shown by one closest to *Trichodorus primitivus*. These are both enoplid nematodes from Clade II, from which relatively few sequences are available, therefore it is not unexpected that no close matches were found. These MOTU, therefore, are likely to have come from enoplid nematodes but cannot be placed to any greater degree of taxonomic resolution based on the data here. When these sequences were searched against the whole Genbank nucleotide dataset no higher-scoring matches were found, so it is unlikely that these sequences are of non-nematode origin.

## 4.7 Phylogenetic Analysis

All MOTU consensus sequences were aligned, together with a set of named sequences from existing databases (see Appendix 1 for details), using ClustalW (Thompson et al. 1994). The resulting alignment was used for phylogenetic analysis in PAUP. The neighbour-joining algorithm was used to create a tree (Figure 4.7.1), with absolute number of nucleotides as a distance measure, 'missing data' sites (gaps and Ns) ignored, and the nematomorph *Gordius aquaticus* as an outgroup. Since the length of sequence used is relatively short and no explicit model of evolutionary change is employed, the branching order should not be considered a rigorous statement of actual evolutionary relationships between taxa: this tree is not intended to represent deep phylogenetic relationships but only the degree of difference between taxa at the tips of the tree.

The MOTU sequences appear to represent a wide range of biological diversity. Representatives are found from across the phylum Nematoda, with taxa from clades I, II, IV and V in the molecular phylogeny (Blaxter et al. 1998). Clade III is entirely animal-parasitic and therefore would not be expected to appear in a survey of free-living nematodes. A notable feature of the tree is the presence of at least two 'taxon clouds'. At the top of the tree as displayed here lies a large set of MOTU sequences close to that determined for *Helicotylenchus*, including the most abundant taxon, 2bp_MOTU0002, but also round 30 other MOTU with far fewer sequences, showing a close similarity to this abundant MOTU (differing by no more than 10 bases), and yet distinct from it by the heuristics employed here. Similarly, a large set of MOTU showing similarity to nematodes within the order Dorylaimida are seen, though in this case sequences group with several different genera, including *Aporcelaimellus*, *Paractinolaimus* and *Eudorylaimus*. It is not clear from this information whether these groups of MOTU represent a real

biological phenomenon (i.e. there genuinely are a large number of similar taxa within the *Helicotylenchus* and dorylaimid groups, distinct by SSU sequence, at the Sourhope field site) or an artefact of the MOTU-estimation process (i.e. this variation reflects only sequencing errors or other methodological problems, in which case these results may significantly overestimate the true number of taxa). The next chapter will endeavour to address this point with a more detailed examination of the variation within these two sets of sequences.

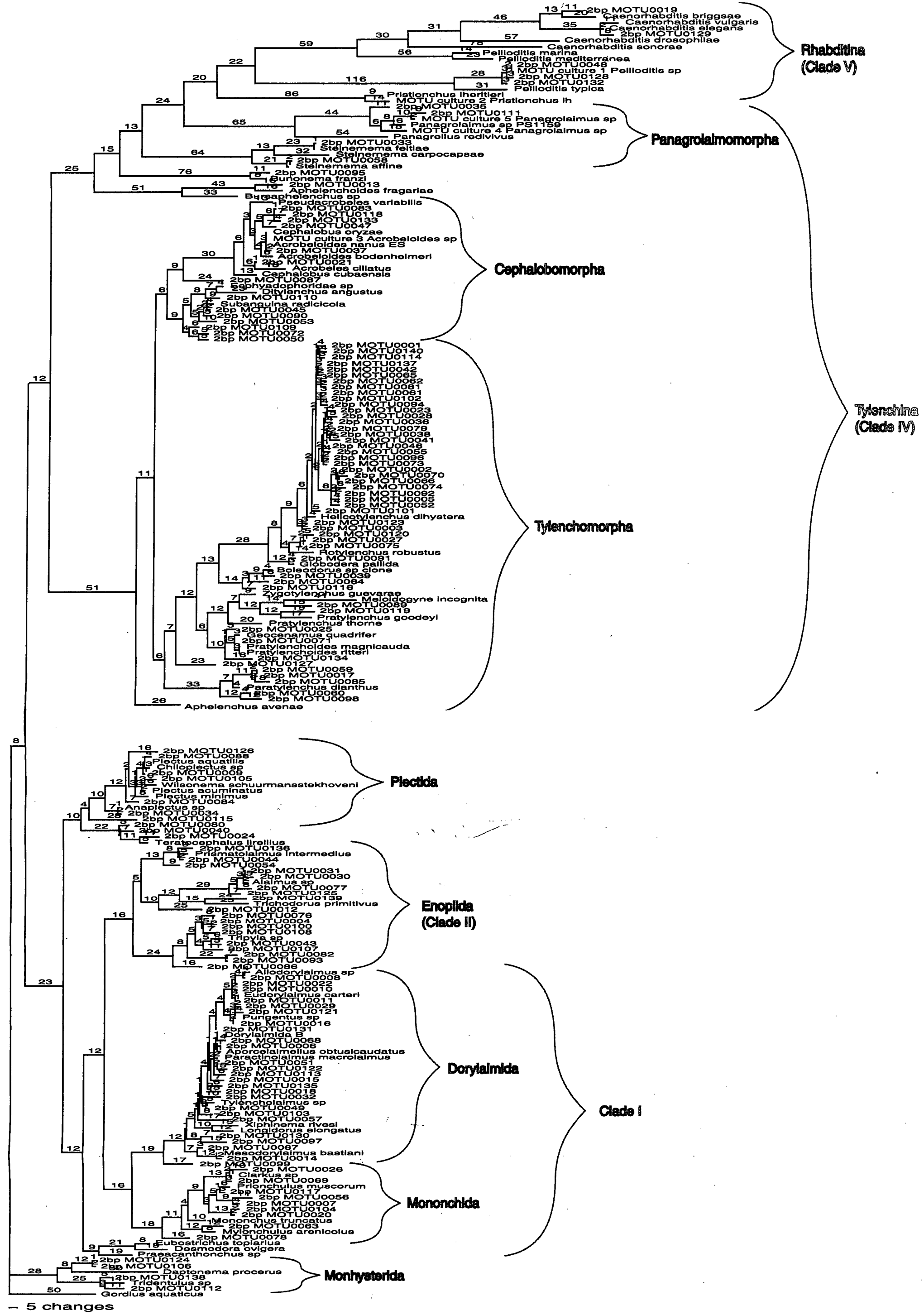



Figure 4.7.1 Phylogram showing a neighbour-joining analysis of MOTU consensus sequences and a selected set of identified nematode sequences, using absolute nucleotide difference as distance measure and rooted using *Gordius aquaticus* as outgroup. Branch lengths are given, indicating numbers of base changes. Clade names are taken from Blaxter et al (1998) and De Ley and Blaxter (2001).

# 5. Testing the Robustness of Molecular Diversity Estimation

Any effort to classify biological variation carries with it particular statistical issues. Common problems include the difficulty of distinguishing natural variation from experimental error, and possible artefactual results produced by the process of classification. This chapter examines the concerns associated with defining taxa on the basis of DNA sequence similarity, and tests the robustness and reliability of the methods employed.

MOTU are defined on the basis of sequence similarity, yet it is known that sequence differences can arise from both natural variation and from errors in the PCR or sequencing process. In any single 475-base sequence, the probability of an error is low. But when hundreds or thousands of such sequences are generated, the number of potential random errors within this dataset becomes considerable. Whatever similarity threshold is chosen for MOTU assignment, the possibility exists that a single sequence may contain a number of errors greater than this threshold, and could thus potentially be misclassified. The likelihood of such a situation increases with the number of sequences generated. Therefore, it is important to know the experimental error rate before interpreting any set of MOTU designations.

Another consideration is the fact that the order of searching of sequences can have an effect on the number of MOTU defined. Consider three hypothetical sequences, A, B and C, where A differs from B by two bases, B differs from C by two bases, but A and C differ by three bases. Suppose these sequences are searched in the order A, B, C, with two base differences set as the threshold similarity for MOTU designation. When B is searched against A, it will be assigned to the same MOTU since it differs by only two bases. C will then be searched against A and B, B will be found as the closest match, and since it again differs by only two bases, C will be assigned to the same MOTU as B. Therefore, all of these sequences will be assigned to a single MOTU.

However, suppose the same sequences are searched in the order A, C, B. When C is compared against A, it will be found to differ by more than two bases, and will therefore be assigned to a new MOTU. When B is searched, it will be arbitrarily assigned to the same MOTU as either A or C (other things being equal, the sequence giving the longer match will be chosen). Thus, simply by searching the same sequences in a different order, a different number of MOTU can be found. It is therefore necesary to test the frequency with which this type of situation occurs in the real data in order to know the degree of confidence we may attach to these classifications.

## 5.1 Effect of Sequencing Errors

### 5.1.1 Measurement of error rate

In order to obtain an empirical estimate of the error rate resulting from PCR and/or sequencing errors, multiple individuals were picked and sequenced from the same set of cultures. Since, in all of these cases, the 'true' sequence is known, any deviations must result from errors. Twelve nematodes were picked from each of the cultures ED2003, ED2012, ED2055, ED2063, ED2074 and ED2086. After trimming of low quality sequence by the standard method, 34,603 bases were generated, which were found to contain a total of 10 errors. This suggests an average per-base error rate of approximately 0.0003; that is, for each base, the probability of an error is 1 in 3460, assuming that the error rate is constant across all sequences and all sites.

On this basis, the probabilities of multiple errors in the same sequence were calculated. If sequencing errors are random, we would expect their distribution to follow a curve decreasing away from zero – i.e. a certain fraction of sequences should contain no errors, a smaller fraction should contain one error, a still smaller fraction should contain two errors, and so on. The mean sequence length was 475 bases. Therefore the probaility of a single error in a 475 base sequence is 0.0003 x 475 = 0.137. The probability of two errors in the same sequence is given by 0.137 x (0.0003 x 474) = 0.019, the probability of an error in the remaining 474 bases given that an error has occurred once. Thus the probaility of three errors is 0.019 x (0.0003 x 473) = 0.003, and so on. Each of these probabilities was then multiplied by 2039 to determine the number of sequences we would expect to contain this number of errors in a sample of 2039 (see Table 5.1.1).

| No. errors | Probability per 475bp seq. | No. exp. in 2039 seqs. |
|---|---|---|
| | | |
| 0 | 0.841 | 1714.695 |
| 1 | 0.137 | 279.896 |
| 2 | 0.019 | 38.341 |
| 3 | 0.003 | 5.241 |
| 4 | 3.506E-04 | 0.715 |
| 5 | 4.772E-05 | 0.097 |
| 6 | 6.482E-06 | 0.013 |
| 7 | 8.786E-07 | 0.002 |
| 8 | 1.188E-07 | 2.423E-04 |
| 9 | 1.604E-08 | 3.270E-05 |
| 10 | 2.160E-09 | 4.403E-06 |

Table 5.1.1 Probabilities of increasing numbers of errors based on observed error rate from culture resequencing, and the consequent expected numbers in the 2039 sequences generated in this survey.

These calculations suggest that the majority (~99.7%) of sequences will contain 0, 1 or 2 errors. The remaining ~0.03% are expected to contain 3 or more errors. Thus a difference of 2 bases or less would appear a reasonable minimum level of similarity to define MOTU. For the 2039 sequences sampled by this survey, this would correspond to approximately 2033 sequences containing 0, 1 or 2 errors, and approximately 6 containing more. Thus the number of MOTU may potentially be overestimated by 6, and this uncertainty should be borne in mind when interpreting subsequent results. However, sequences containing 3 or more errors will not automatically be placed in the wrong MOTU. There may be an intermediate sequence which is able to 'pull' it into the correct cluster; or, if a MOTU is a singleton and is genuinely distant from any other sequence, even multiple errors will not prevent it from remaining a singleton.

However, as already stated, the above reasoning depends crucially on the assumption that the error rate is constant across all sequences and all sites. In reality, there are reasons to believe that this is not always the case: certain sequences are likely to be of lower overall quality than others, and hence more likely to contain multiple errors than would be suggested by simply calculating an average error rate and applying it across all sequences. Additionally, in the culture sequences tested, it was observed that while most error-containing sequences contained a single error, one contained three errors, all adjacent to one another. An examination of the trace file made it clear that these three errors were due to an artefactual guanine peak which had obscured the true bases across these three sites, probably due to a sequencing reaction that had not been fully cleaned. This demonstrates that certain types of 'error-causing event' can produce more than one base difference in a single sequence.

### 5.1.2 A test set: the *Helicotylenchus* "flock"

The set of sequences close to the SSU of *Helicotylenchus* sp. (see Chapter 4) is a group which may be considered 'problematic' in terms of reliably resolving into taxa – it contains much variation, but due to the sheer number of sequences in the group, some (or perhaps even all) of this variation may be due to errors. It is possible that several distinct sequence types exist, but also possible that only one sequence (the most common type) is 'real', and all other MOTU merely represent error-containing copies of this sequence. Therefore, all sequences whose closest match by BLAST analysis was *Helicotylenchus* (877 in total) were reanalysed separately as a test set to determine the robustness of the MOTU-estimation process.

The level of variation within this set of sequences was examined using the **base_diff.pl** script (see Appendix 3). This time, only one sequence was used as a BLAST database: 17474ED was chosen as a representative of the most common type. Each of the 877 sequences was compared to this single sequence, to determine the number of bases by which they differed. A summary of results is shown in Table 5.1.2, together with the number or sequences expected to contain the corresponding number of errors, based on multiplying the error probailities calculated in 5.1.1 by 877. Both sets of numbers are plotted in Figure 5.2.2.

Differences between the sequence being tested and the reference sequence (17474ED) may derive from two sources: they may represent 'real' , natural variation in the sequences of these nematodes, or they may be due to experimental errors; it is likely that both are contributing. The overall shape of the graph of observed differences in Figure 5.1.2 appears similar to that calculated from the error rate in section 5.1.1, though greater than expected numbers are seen for the larger differences. This may reflect the fact that poor quality sequences are likely to contain multiple errors, but may also indicate that there is real variation within this set. In particular, the observed data appears to diverge from expected values at two bases and four bases from the reference sequence, which may indicate that there are genuine OTU differing by these numbers of bases from the most common type.

| No. base differences | No. sequences | No. exp. in 877 seqs. |
|---|---|---|
| | | |
| 0 | 667 | 737.512 |
| 1 | 102 | 120.387 |
| 2 | 52 | 16.491 |
| 3 | 20 | 2.254 |
| 4 | 19 | 0.307 |
| 5 | 6 | 0.042 |
| 6 | 3 | 5.685E-03 |
| 7 | 4 | 7.705E-04 |
| 8 | 1 | 1.042E-04 |
| 9 | 1 | 1.406E-05 |
| 10 | 2 | 1.894E-06 |

Table 5.1.2 Numbers of bases by which each *Helicotylenchus*-like sequence differs from a single representative (17474ED), and the expected numbers of errors for 877 sequences based on the error rate calculated in section 5.1.1 and assuming only one 'true' sequence.

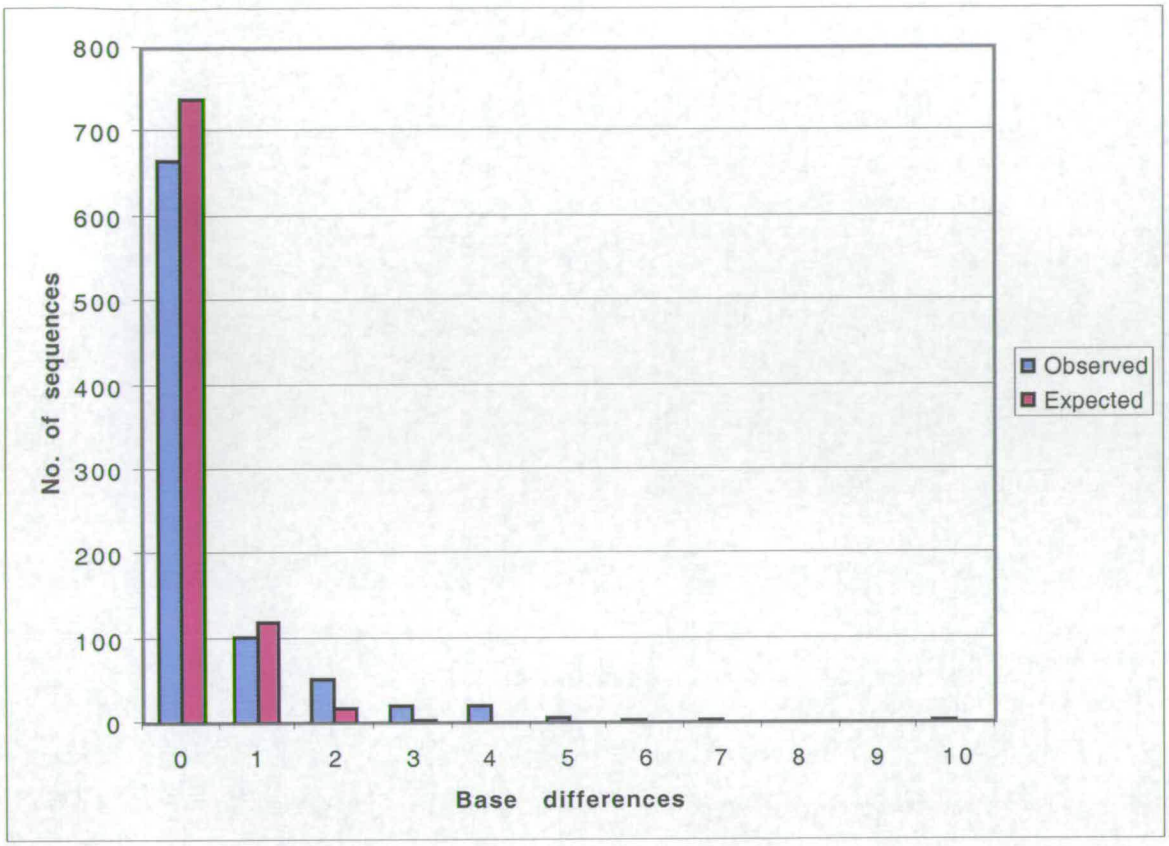



Figure 5.1.2 Graph of number or observed base differences among the *Helicotylenchus*-like sequences plotted alongside the expected numbers of errors.

One route by which genuine taxa might be distinguished is by determining the number of times each MOTU occurs: if the same sequence variant is found to occur more than once, it is unlikely that this variant is due to a random error. To examine this, **define_MOTU.pl** was used to produce a set of '0bp_MOTU' from these sequences (i.e. only 100% identical sequences placed in the same MOTU, with no variation allowed other than gaps or Ns). As shown in Table 5.1.3, each of the sequences differing by 5-10 bases from 17474ED is sampled only once (i.e. no two sequences cluster into the same MOTU, therefore the number of MOTU in a category is the same as the number of sequences). However, for one, two and four base differences, the number of MOTU is fewer than the number of sequences, meaning that certain sequences were identical to each other.

The paradoxical result that three 0bp_MOTU are found for sequences which are identical to the reference sequence is caused by two short sequences, which are indeed identical to 17474ED over the region compared, but are also identical to other sequences already assigned to different MOTU because of variations at one end or the other, information which is missing in the short sequences. Thus, in these two instances, the short sequences have been arbitrarily assigned to different MOTU, because more than one MOTU produced an equally high BLAST score. This illustrates the difficulty caused by variations in sequence length, which could only be entirely eliminated if all sequences were exactly the same length.

There are 25 distinct MOTU which are sampled more than once; since they occur repeatedly these variations are unlikely to be due to random errors. It is interesting to note that all but two of these differ from the reference sequence by only one or two bases, and therefore by normal heuristics would all be placed in the same MOTU. This suggests the possibility that some degree of real sequence variation will be 'lumped' together by choosing two bases as the MOTU designation threshold. It is unavoidable that, whatever level is chosen, there is a trade-off between resolution and reliability – choosing a lower number of bases enables greater sensitivity to variation, but can also result in a greater number of MOTU being erroneously defined, while choosing a higher number may cause real variation to be overlooked but increases the confidence associated with the MOTU which are defined.

| No. base diffs. | No. sequences | No. 0bp_MOTU | MOTU with >1 seq. |
|---|---|---|---|
| | | | |
| 0 | 667 | 3 | 1 |
| 1 | 102 | 54 | 17 |
| 2 | 52 | 44 | 5 |
| 3 | 20 | 20 | 0 |
| 4 | 19 | 17 | 2 |
| 5 | 6 | 6 | 0 |
| 6 | 3 | 3 | 0 |
| 7 | 4 | 4 | 0 |
| 8 | 1 | 1 | 0 |
| 9 | 1 | 1 | 0 |
| 10 | 2 | 2 | 0 |
| | | | |
| Totals | 877 | 150 | 25 |

Table 5.1.3 Numbers of 0bp_MOTU represented by the *Helicotylenchus*-like sequences in each base difference category, and the numbers of MOTU containing more than one sequence.

### 5.1.3 A second test set: Dorylaimida

The tree of all MOTU also shows a 'cloud' of dorylaimid-like sequences. This is the largest group of sequences in the dataset after the *Helicotylenchus* group, and so all of the dorylaimid sequences were also separated and analysed apart from the rest. All sequences for which the top BLAST hit was a member of the order Dorylaimida was extracted from the SQL database (666 sequences in total). Although they still represent a well-defined clade within the tree, these sequences are expected to be a more diverse group than the *Helicotylenchus*-like set, since sequences showed similarity to a number of different genera rather than a single genus.

As before, a single sequence representing the most common MOTU (15718ED, closest to *Aporcelaimellus obtusicaudatus*) was selected and all other sequences were compared against this one using BLAST to determine the number of base differences. A summary of the results is shown in Table 5.1.4, and plotted as a graph in Figure 5.1.4.

| No. base diffs. | No. sequences | No. 0bp_MOTU | MOTU with >1 seq. |
|---|---|---|---|
| | | | |
| 0 | 164 | 2 | 1 |
| 1 | 127 | 19 | 5 |
| 2 | 25 | 13 | 3 |
| 3 | 3 | 3 | 0 |
| 4 | 8 | 3 | 1 |
| 5 | 0 | 0 | 0 |
| 6 | 12 | 2 | 1 |
| 7 | 8 | 6 | 1 |
| 8 | 128 | 8 | 5 |
| 9 | 36 | 19 | 8 |
| 10 | 16 | 10 | 1 |
| 11 | 110 | 5 | 4 |
| 12 | 17 | 9 | 1 |
| 13 | 0 | 0 | 0 |
| 14 | 3 | 3 | 0 |
| 15 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 |
| 17 | 3 | 2 | 1 |
| 18 | 1 | 1 | 0 |
| 19 | 0 | 0 | 0 |
| 20 | 1 | 1 | 0 |
| 21 | 0 | 0 | 0 |
| 22 | 3 | 2 | 1 |
| 23 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 |
| 26 | 0 | 0 | 0 |
| 27 | 0 | 0 | 0 |
| 28 | 0 | 0 | 0 |
| 29 | 0 | 0 | 0 |
| 30 | 1 | 1 | 0 |
| | | | |
| Totals | 666 | 109 | 33 |

Table 5.1.4 Numbers of 0bp_MOTU represented by the Dorylaimid sequences in each base difference category, and the numbers of MOTU containing more than one sequence.

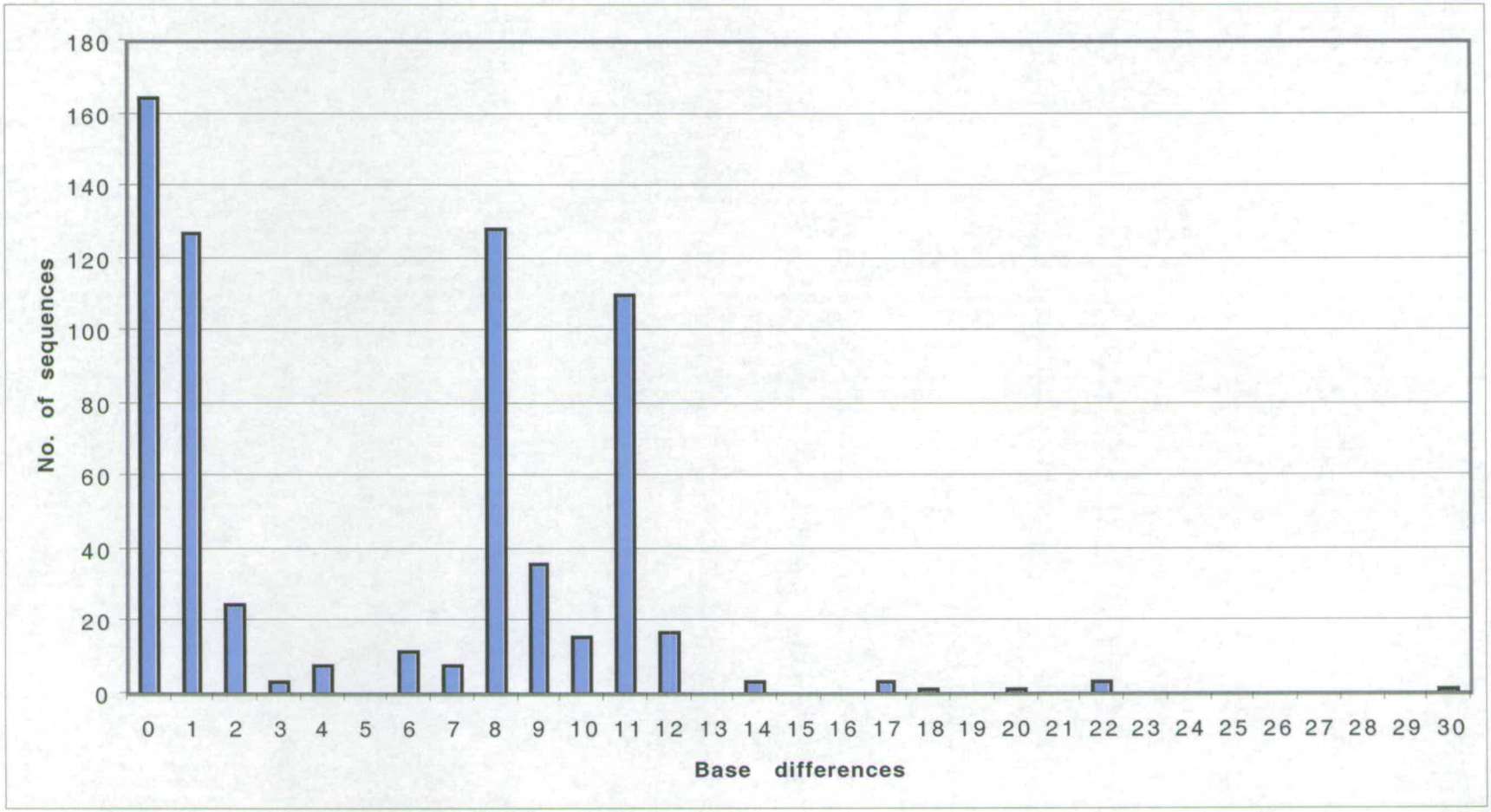





Figure 5.1.4 Numbers of bases by which each dorymid sequence differs from a single representative (15718ED).

From this graph (Figure 5.1.4) it can be seen that there are at least three distinct peaks, at 0, 8 and 11 bases from the reference sequence, showing that these numbers of differences are found repeatedly. This suggests that there are at least three 'real' taxa distinguishable by sequence, which are found repeatedly and are therefore not likely to result from errors. The distribution of variation becomes more complex: the fact that there are three taxa with large numbers of members (over 100) means that the errors are now distributed around at least three 'standards' instead of just one. A closer examination suggests that there may be even more than this: there are smaller peaks at 4, 6, 17 and 22 bases, as well as a far higher number of sequences at 1 base difference than would be expected at random. Overall, there are a total of 33 0bp_MOTU which are represented by more than one sequence, and are therefore likely to be real. As expected, these results reveal the set of Dorylaimid sequences to be more heterogeneous than the *Helicotylenchus* group, with more clearly defined MOTU.

## 5.2 Distribution of variation by position across the sequence

### 5.2.1 Location of variable sites

Another route toward separating real variation from errors is to examine where the variation occurs in the molecule. Errors are expected to be randomly spaced; real differences (but also non-random errors, should they occur) are likely to be concentrated in certain variable sites. To examine the distribution of variation, an alignment of all 877 sequences in the *Helicotylenchus* set was constructed using the ClustalX software. The alignment produced was 486 characters in length, but all columns containing predominantly gaps were excluded, leaving 475 columns. The variation at each site in the sequence was analysed: for each column in the alignment, the consensus (most common) base and the variability (the number of positions in the column which differ from this consensus, excluding gaps or Ns) were determined.

Results are shown as a histogram in Figure 5.2.1 (a), the upper of the two graphs. Overall it was found that 361 out of the 475 sites (76%) had no variation; the remaining sites ranged from 1 base to 40 bases variation within one site, out of a total of 877 sequences (i.e. 877 characters for each aligned site). From these results it can be seen that certain sites within the molecule are more variable than others; there is also variation of the type which would be expected from random sequencing errors, with sites showing one or two bases variation distributed across the molecule seemingly at random. Again, the problem arises of distinguishing between which differences are 'real' and which are due to experimental error. It is necessary to define a level of variability above which a site is considered 'significantly' more variable than would be expected at random.

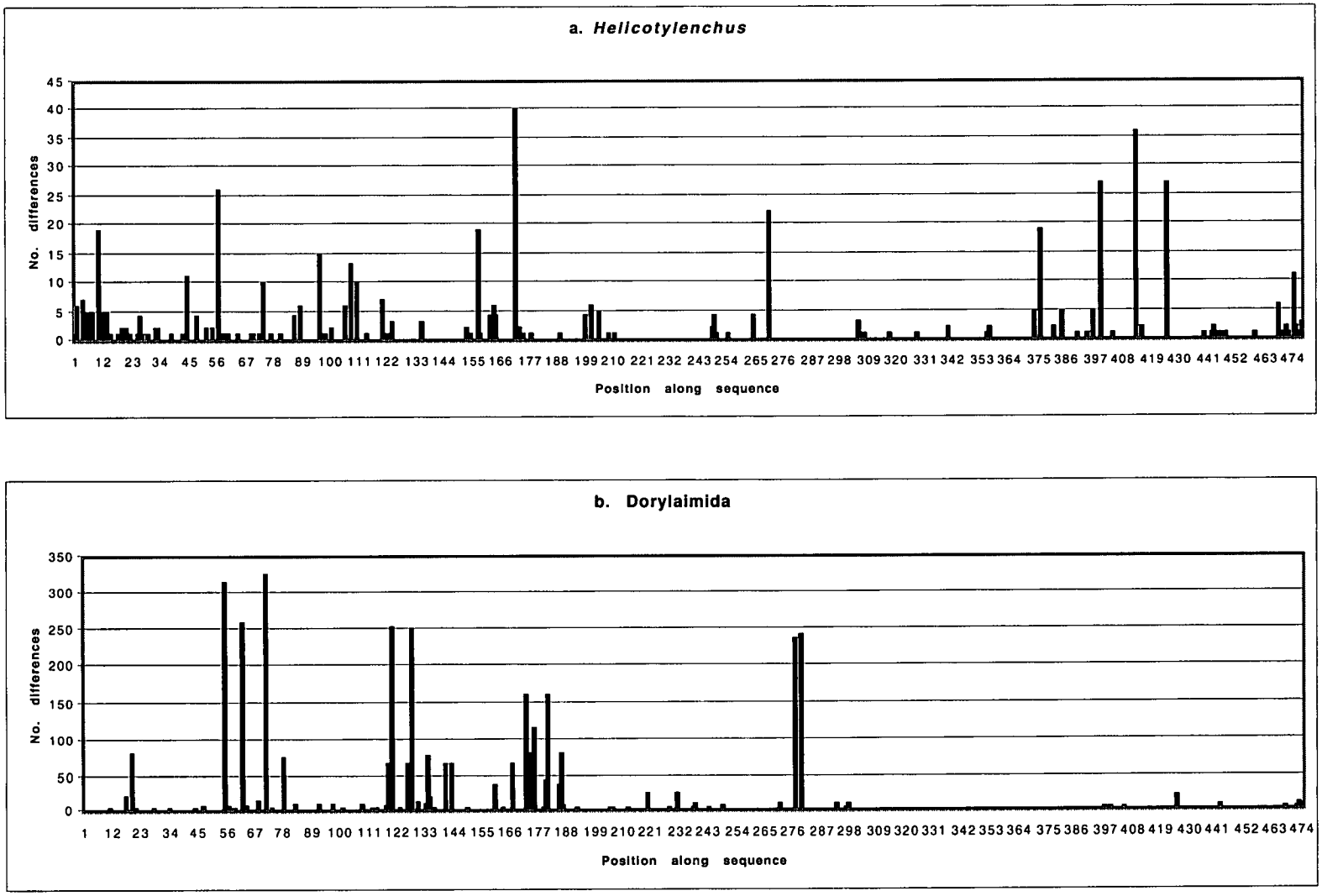



Figure 5.2.1 Map of sequence variation across an alignment of 475 nucleotides for (a) the 877 *Helicotylenchus*-like sequences, and (b) the 666 dorylaimid sequences (y-axis=number of bases differing from the consensus base for that site).

Since an estimate of the average per-base error rate is already available, it is possible to estimate the distribution of expected random errors by aligned site, as was done before by sequence. If the error rate is 0.0003 per base, and this rate is assumed to be constant, the probability of a single error within a column of 877 bases is 0.0003 x 877 = 0.253; the probability of two errors in a column is 0.253 x (0.0003 x 876) = 0.064, and so on. The actual numbers expected in each category can be calculated by multiplying by the length of the alignment (475 characters for the current example). These expected numbers are listed in Table 5.2.1.

This provides a set of numbers, expected under a random model, against which the observed data from the *Helicotylenchus* set can be compared. The observed data appear to deviate from expectations at several points. There are more invariant sites (361 as opposed to 314), and fewer sites with a single difference than would be expected. Sites differing by 5 bases also appears to show a small but distinct peak in frequency. Also, the probability of a site containing more than 5 random errors is small, and becomes even smaller for greater numbers of errors, yet sites are observed with as many as 40 differences. This suggests that, although random errors are undoubtedly contributing to this variation, there is nevertheless variation which does not fit the expectations of this simple random model.

| No. errors | Prob. per aligned site | No. exp. in 475 aligned sites |
|---|---|---|
| | | |
| 0 | 0.661 | 313.827 |
| 1 | 0.253 | 120.387 |
| 2 | 0.064 | 30.477 |
| 3 | 0.016 | 7.707 |
| 4 | 0.004 | 1.947 |
| 5 | 0.001 | 0.491 |
| 6 | 2.605E-04 | 0.124 |
| 7 | 6.558E-05 | 0.031 |
| 8 | 1.649E-05 | 7.832E-03 |
| 9 | 4.141E-06 | 1.967E-03 |
| 10 | 1.039E-06 | 4.934E-04 |
| 11 | 2.603E-07 | 1.236E-04 |
| 12 | 6.513E-08 | 3.094E-05 |

Table 5.2.1 Probabilities of increasing numbers of errors per column based on the error rate calculated in section 5.1.1, and the numbers expected in 475 columns

| Base diffs. | Expected | Observed |
|---|---|---|
| | | |
| 0 | 313.827 | 361 |
| 1 | 120.387 | 52 |
| 2 | 30.477 | 17 |
| 3 | 7.707 | 3 |
| 4 | 1.947 | 8 |
| 5 | 0.491 | 11 |
| 6 | 0.124 | 6 |
| 7 | 0.031 | 2 |
| 8 | 7.832E-03 | 0 |
| 9 | 1.967E-03 | 0 |
| 10 | 4.934E-04 | 2 |
| 11 | 1.236E-04 | 2 |
| 12 | 3.094E-05 | 0 |
| 13 | 7.734E-06 | 1 |
| 14 | 1.931E-06 | 0 |
| 15 | 4.816E-07 | 1 |
| 16 | 1.200E-07 | 0 |
| 17 | 2.985E-08 | 0 |
| 18 | 7.419E-09 | 0 |
| 19 | 1.842E-09 | 3 |
| 20 | 4.567E-10 | 0 |
| 21 | 1.131E-10 | 0 |
| 22 | 2.798E-11 | 1 |
| 23 | 6.913E-12 | 0 |
| 24 | 1.706E-12 | 0 |
| 25 | 4.206E-13 | 0 |
| 26 | 1.036E-13 | 1 |
| 27 | 2.547E-14 | 2 |
| 28 | 6.256E-15 | 0 |
| 29 | 1.535E-15 | 0 |
| 30 | 3.762E-16 | 0 |
| 31 | 9.208E-17 | 0 |
| 32 | 2.251E-17 | 0 |
| 33 | 5.497E-18 | 0 |
| 34 | 1.341E-18 | 0 |
| 35 | 3.267E-19 | 0 |
| 36 | 7.949E-20 | 1 |
| 37 | 1.932E-20 | 0 |
| 38 | 4.690E-21 | 0 |
| 39 | 1.137E-21 | 0 |
| 40 | 2.754E-22 | 1 |

Table 5.2.2 Base differences per aligned site, expected based on the calculated error rate, and observed in the *Helicotylenchus* dataset

Based on the frequencies calculated above, we would expect approximately 99.9% of columns to contain between 0 and 4 errors; therefore, any site showing 5 or more differences may be considered significantly variable at the p=0.001 level. On this basis, in the *Helicotylenchus* dataset of 877 sequences, 34 sites out of 475 are significantly variable. However, the actual number of MOTU implied by this number is smaller, since it is known that at least some of this variation is linked within the same sequences (see Table 5.1.2), with as many as 10 differences sometimes seen within the same sequence. A number of MOTU in the 20-30 range, for example, would be close to the estimate of 25 arrived at by counting the recurrence of each 0bp_MOTU (see Table 5.1.3). It would therefore be advantageous to further examine the pattern of linkage between different variable sites.

For comparison, the same analysis was run on an alignment of the 666 dorylaimid sequences; a map of the variability along the sequence length is shown as Figure 5.2.1 (b). As expected, the dorylaimid set - even though it contains fewer members - shows greater variation than was seen for the *Helicotylenchus* sequences. An immediately noticeable difference is the scale on the y-axis – while the most variable site in the *Helicotylenchus* set contained 40 variants, the maximum for the dorylaimid set is 324, and there are a further 20 sites whose variability is above 40. The number of sites showing 5 differences or greater (the criterion established for significantly variable sites) is 49. These findings are consistent with previous results suggesting that the dorylaimid sequences are a more heterogeneous group containing several clearly distinct OTUs.

### 5.2.2 Linkage of variation

To examine the question of whether variations are seen in the same sequences across different aligned sites, a PERL script was written which took each column in turn, and determined which sequences differed from the consensus base. It then reanalysed the entire dataset with those sequences temporarily removed from every column, and calculated the new variability scores for each aligned site. This procedure was repeated for every column which contained variation, and for each column a number was obtained indicating how much overall variability had been removed by eliminating the variable sequences in that column from the entire set. If the number of sequences removed and the reduction in variation are the same, this indicates that those sequences are not linked to variation at any other site. If the two numbers differ, this shows that those sequences are linked to variation at other sites.

Results for the *Helicotylenchus* set are plotted in Figure 5.2.2. This shows a graph in which each point represents the result of removing all the variable sequences at a particular aligned site from the entire dataset. For example, the point at 15 on the x-axis and 35 on the y-axis means that in this particular column, 15 sequences were variable, and that after these 15 sequences were removed from all columns, overall variation decreased by 35, showing that those 15 sequences were also variable at a number of other sites. The line shows the pattern which would be expected if variation were entirely unlinked – for example, removing 20 sequences would always result in overall variation decreasing by just 20, as no other column would be affected by the removal of the 20 variable sequences in that column. Therefore, the further a point is from this line, the greater the degree to which the variable sequences at that site are also variable at other sites.

It was found that 104 out of the 114 variable sites are linked to other sites, in that the decrease in overall variability was greater than the number of sequences removed; some of this association between sites is likely to be due to random chance. For 84 of these sites the decrease in variability is more than double the number of sequences removed. It therefore appears that sequences which are variable at a particular site are often variable at other sites also.

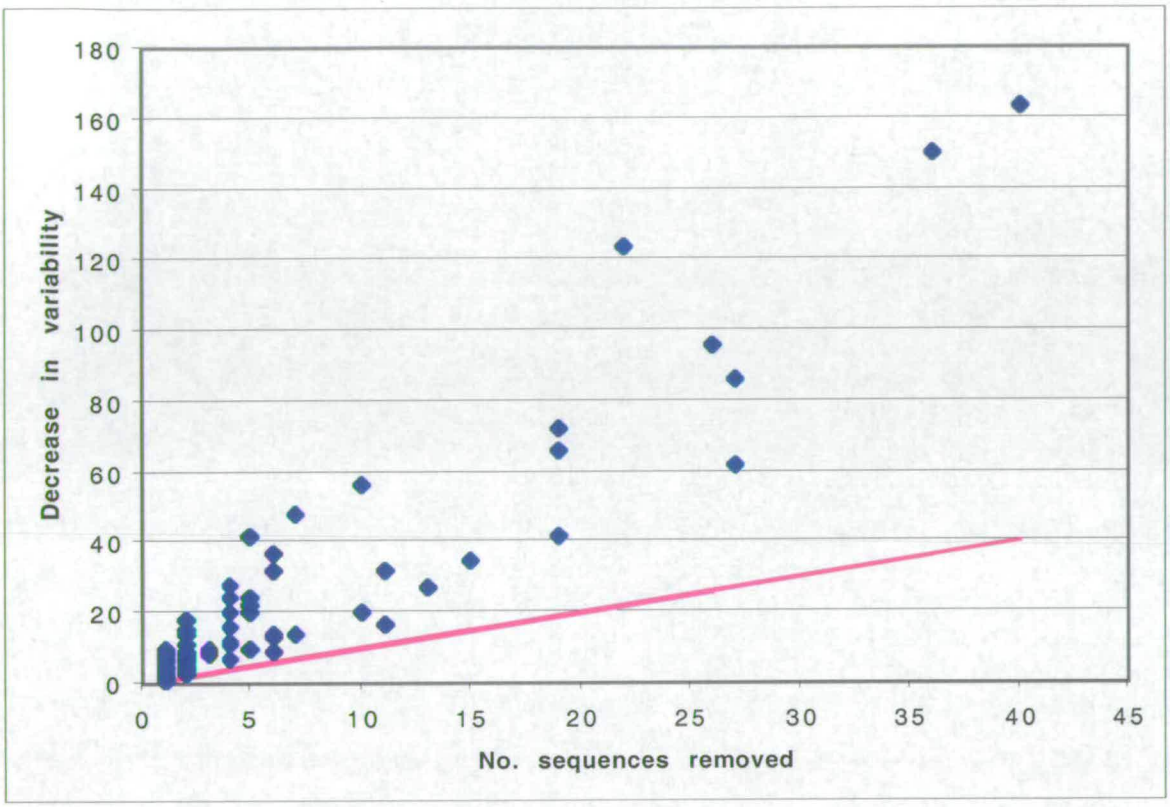



Figure 5.2.2 Graph showing linkage between variation at different sites within the *Helicotylenchus* sequence set. Each point represents the result of removing all the variable sequences at a particular aligned site from the entire dataset. The line shows the pattern which would be expected if variation were entirely unlinked.

### 5.2.3 Sequence variability at significant sites

Having examined the distribution of variation within an alignment both by sequence (rows) and by site (columns), it is possible to combine the two by determining, from a set of sequences containing one or more differences from the dominant type, how many of those differences are located at significantly variable sites. From the *Helicotylenchus* data analysed in section 5.1.2, the position of each variation relative to the beginning of the 17474ED sequence was calculated, compared against the set of significantly variable sites estimated in section 5.2 (those containing 5 or more variants: these are sites 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 44, 56, 74, 88, 96, 106, 108, 110, 120, 157, 163, 172, 201, 204, 270, 373, 376, 384, 396, 399, 413, 425, 468, and 474) and for each sequence, the number of variations found at significantly variable sites was determined. A summary of results, together with previous figures for comparison, is shown in Table 5.2.3 below.

| No. base differences | No. sequences – all sites | No. sequences – variable sites only |
|---|---|---|
| | | |
| 0 | 667 | 688 |
| 1 | 102 | 109 |
| 2 | 52 | 40 |
| 3 | 20 | 20 |
| 4 | 19 | 8 |
| 5 | 6 | 6 |
| 6 | 3 | 4 |
| 7 | 4 | 2 |
| 8 | 1 | 0 |
| 9 | 1 | 0 |
| 10 | 2 | 0 |

Table 5.2.3 Numbers of bases by which each *Helicotylenchus*-like sequence differs from a single representative (17474ED), with all sites included (middle column), and with only significantly variable sites included (right column).

It can be seen that there is a decrease in overall variation if only significant sites are counted. However, much variation remains, for example the same number of sequences show 3 differences and 5 differences after reanalysis. There are a total of 84 unique sequence types if only significant sites are counted; of these, 27 are sampled more than once. On this basis, therefore, we may confidently say that these 877 *Helicotylenchus*-like sequences contain at least 27 distinct OTUs which are robust both in the sense that they vary only at significantly variable sites, and are represented by more than one sequence. This variation many reflect the fact that, like many soil nematodes, *Helicotylenchus* reproduces mainly as a self-fertilising hermaphrodite. Therefore we would expect a large number of essentially asexual lineages with no genetic exchange between them, which over time could slowly accumulate genetic changes.

## 5.3 Effect of processing order on MOTU assignment

The order in which sequences are processed into MOTU is expected to have some effect on the set of MOTU defined. The importance of this effect in the real data can be empirically tested by running **define_MOTU.pl** multiple times on the same set of sequences, with sequences provided in a random order each time, and analysing the similarities and differences in the MOTU produced.

The *Helicotylenchus*-like sequences were again chosen as a test set. These 877 sequences were processed into 2bp_MOTU 100 times, with sequence order randomised for each trial. The number of MOTU defined ranged from 21 to 30, with a mean of $25.15 \pm 2.042$. The same analysis was carried out on the 666 Dorylaimid sequences; here the number of MOTU ranged from 21 to 28, with a mean of $23.55 \pm 1.41$. Both set of results are summarised in Table 5.3.1 below, and plotted as a histogram in Figure 5.3.1.

This shows that, as expected, running the same set of sequences multiple times will give different numbers of MOTU depending on the order in which the sequences are processed. From examining the set of MOTU assignments for the *Helicotylenchus*-like sequences, the only MOTU which are consistent in every run are the single member MOTU. There is always one very large MOTU containing >800 sequences, but the precise number of sequences assigned to this large MOTU varies between analyses. There are 39 sequences differing by three or four bases from the standard consensus sequence, which are sometimes assigned to the large MOTU and sometimes not, depending on the order of searching and whether or not a sequence of intermediate similarity is able to 'pull' the more divergent sequence into the same MOTU.

The Dorylaimida set shows somewhat less variation in the number of MOTU defined by different runs. This may simply reflect the fact that it contains fewer sequences, but also that there are known to be a number of distinct MOTU with significant numbers of members within this group, therefore there are fewer 'borderline' sequences which jump from one MOTU to another in different analyses.

These results suggest that there is uncertainty in the number of MOTU present when a large number of sequences is analysed, and this should be taken into account when interpreting any set of MOTU assignments.

| No. MOTU | Helicotylenchus | Dorylaimida |
|---|---|---|
| | | |
| 21 | 1 | 4 |
| 22 | 10 | 18 |
| 23 | 13 | 33 |
| 24 | 14 | 24 |
| 25 | 19 | 10 |
| 26 | 18 | 8 |
| 27 | 12 | 2 |
| 28 | 7 | 1 |
| 29 | 4 | 0 |
| 30 | 2 | 0 |
| 30+ | 0 | 0 |
| | | |
| Mean | 25.15 | 23.55 |
| St. dev. | 2.042 | 1.410 |

Table 5.3.1 Numbers of MOTU defined by 100 runs of each of the *Helicotylenchus* and dorylaimid datasets.

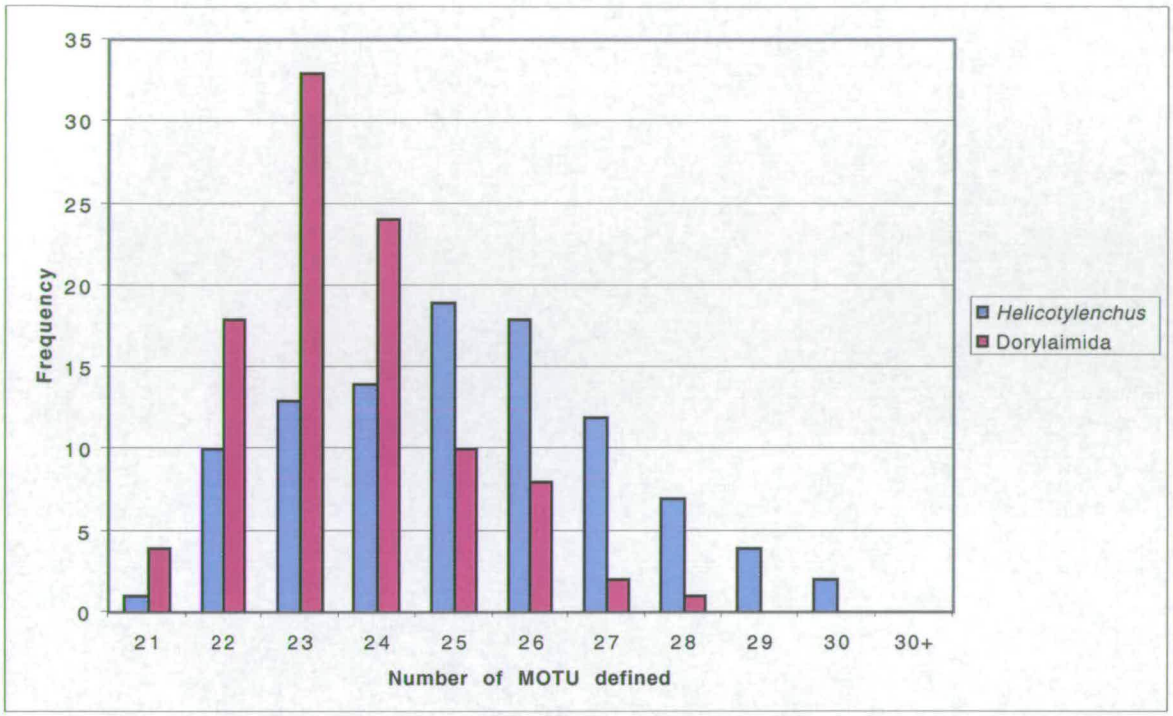



Figure 5.3.1 Numbers of MOTU defined by 100 runs of each of the *Helicotylenchus* and dorylaimid datasets.

## 5.4 Reanalysis of the entire dataset

The above findings suggested that it would be useful to carry out a similar number of runs on the entire dataset, to obtain an estimate of the uncertainty in the overall number of MOTU identified at the site. 100 runs were carried out on all 2039 sequences; results are summarised in Table 5.4.1 below, and plotted as a histogram in Figure 5.4.1.

The number of MOTU varies from 129 to 143 (a range of 15), with a mean of 136.14 ± 2.95. This range is less than the sum of the ranges of the *Helicotylenchus* and dorylaimid sets (10 and 8 respectively), suggesting that most of the variation in MOTU numbers is accounted for by these two large groups, with relatively little variation among the rest of the sequences.

| No. MOTU | frequency |
|---|---|
| | |
| 129 | 2 |
| 130 | 1 |
| 131 | 4 |
| 132 | 4 |
| 133 | 10 |
| 134 | 7 |
| 135 | 11 |
| 136 | 15 |
| 137 | 10 |
| 138 | 12 |
| 139 | 14 |
| 140 | 5 |
| 141 | 2 |
| 142 | 2 |
| 143 | 1 |
| | |
| Mean | 136.14 |
| St. dev. | 2.947 |

Table 5.4.1 Numbers of MOTU defined by 100 runs of the entire dataset (2039 sequences)

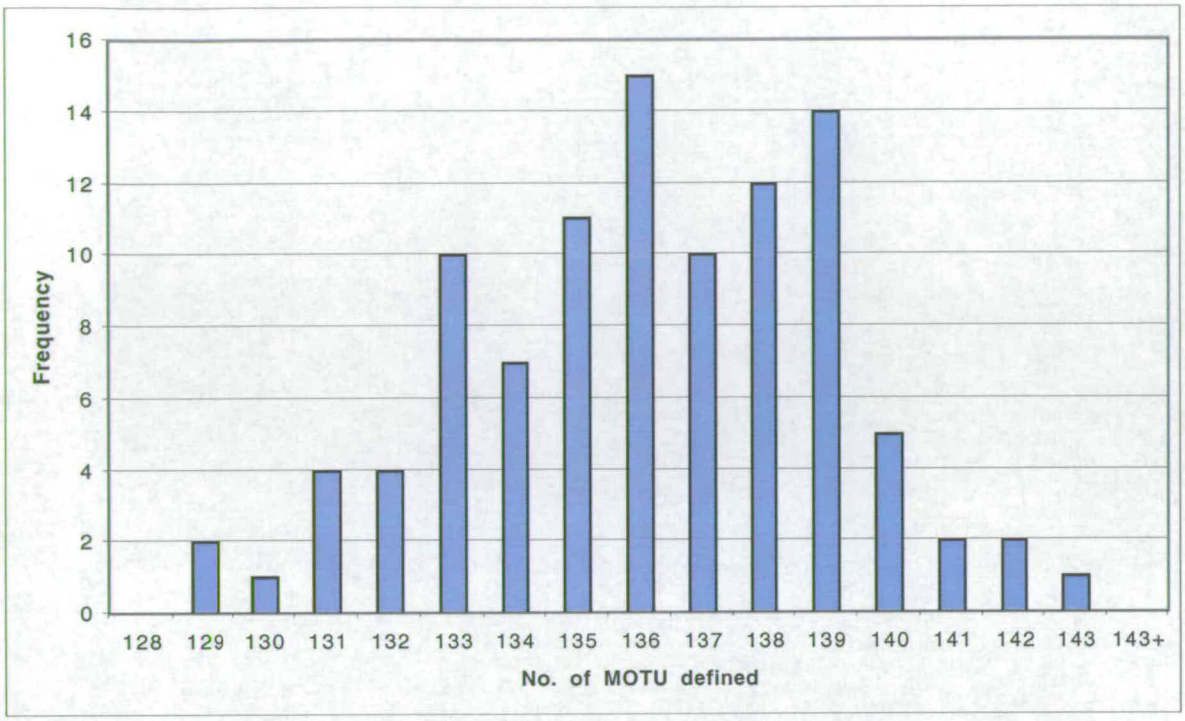



Figure 5.4.1 Numbers of MOTU defined by 100 runs of the entire dataset (2039 sequences)

## 5.5 Conclusions

These findings demonstrate that clustering a set of sequences into OTUs is not as straightforward a matter as it might first appear. Particularly when the number of sequences grows large, numerous sources of uncertainty and variability come into play. It has been shown that the same set of sequences can define different sets of taxa based only on the order in which they are analysed. But it has also been shown that there are methods of analysis which can deal with such uncertainty, and that even very large groups of sequences, within which the observed variation undoubtedly contains many sequencing errors, can be examined in such a way as to separate real variation from errors and provide a robust estimate of the true number of taxa they represent.

These results suggest that any report of the number of MOTU determined from a particular sample should be accompanied with a range or some indication of variability. It is not yet clear how variations in the set of taxa defined from one set of sequences will influence the biological inferences drawn from such information: are these differences sufficient to alter the values of numerical measures of diversity, for example, or to change our estimate of an individual nematode's trophic group or higher taxon? The next chapter will examine these questions.

# 6. Measures of Diversity

## 6.1 Re-estimation of $T_{max}$

On the basis of one run of the process, it was possible to obtain an estimate of the number of taxa which had not yet been sampled (Chapter 4) based on the number of observed taxa represented by one and two individuals; in this run, the observed number of taxa was 140, and the maximum number, $T_{max}$, was estimated to be approximately 417. Since we now have 100 runs on the same set of sequences, it is possible to examine how this estimate changes based on the different runs, and thereby provide a range of estimates.

When the Chao formula (see Chapter 4) was applied to the outputs of all 100 runs, the mean estimate of the total number of taxa was 419.69, with a standard deviation of 31.25, a minimum of 342.13 and a maximum of 504.182. As is to be expected, the range of values is wide, and it is not possible to attach any great certainty to our estimate of the true number of unsampled taxa, given that we must make the assumption that all of the nematode taxon abundances are distributed according to Chao's model (Chao 1984); nevertheless, considering that the maximum number of MOTU actually observed was 143, at the very least these values strongly suggest that a significant proportion, possibly a majority, of the nematode community remains unsampled in our survey of 2039 individuals.

## 6.2 Diversity indices

### 6.2.1 Background

The relationship between species number and abundance of individuals has two main parameters (Southwood and Henderson 2000): (1) taxonomic richness, i.e. number of taxa present; and (2) equitability, or evenness, i.e. the pattern of distribution of individuals among these taxa. For simplicity of comparison of overall diversity between different samples, a common approach is to calculate a single number, referred to as a diversity index, designed to summarise both of these attributes. Over the years a large number of methods for calulating such indices have been proposed, with much debate over the relative merits of different approaches (reviewed by Magurran 1988).

One of the earliest and still most widely used of the diversity indices is the Shannon-Wiener index (*H*) (Shannon 1948), originally devised to determine the amount of information in a code. It is defined:

$$H = -\Sigma p_i (\ln p_i)$$

where $p_i$ is the proportion of individuals in the *i*th species.

Also commonly used is Simpson's index of concentration (Simpson 1949), whose basic form is:

$$C = \Sigma(n_i^2 / N^2) \text{ or } \Sigma p_i^2$$

where $n_i$ is the number of individuals in the $i$th species, and $N$ is the total number of individuals, or $\Sigma n_i$. This index represents the probability that two randomly-chosen individuals from a community will belong to the same species, and ranges from 1/N (maximum diversity) to 1 (no diversity, i.e. all individuals belong to one species). However, this form of the index is not independent of sample size. When sample sizes vary, as in this study, the values of $C$ will not be comparable between samples. However, Simpson also proposed a modification of the index to make it sample size-independent:

$$SI = \Sigma ((n_i^2 - n_i)/(N^2 - N))$$

Like $C$, SI increases as diversity decreases, toward a maximum of 1, representing no diversity. For ease of interpretation it is convenient to use an index that increases as diversity does, which can be done by taking the negative logarithm. Therefore, -ln(SI) provides an index which reflects underlying diversity independently of sample size (Rosenzweig 1995). The Shannon index is more sensitive to changes in absolute number of taxa, while the Simpson index is more dependedent on equitability (Magurran 1988) - it is most strongly driven by the proportionate abundance of the dominant (i.e. most abundant) species.

An index which is simple yet often informative (May 1975) is the Berger-Parker dominance index, $d$ (Berger and Parker 1970) calculated as:

$$d = n_{max} / N$$

where $n_{max}$ is the number of individuals in the dominant species, and $N$ is again the total number of individuals.

Though all of these indices are defined in terms of species, it should in principle be possible to apply them to any kind of taxa, such as MOTU. Indeed, MOTU have properties which make them useful for examining the behaviour of these diversity indices; since multiple runs of MOTU assignment have been carried out on the same set of sequences, it is possible to take the data from all of these runs and examine whether differences in the clustering of sequences result in differences in observed diversity per sample. Furthermore, since MOTU are defined on the basis of exact and objective criteria (a specified number of base differences between each pair of sequences compared) we may examine the effect of varying these criteria on the behaviour of the various diversity indices.

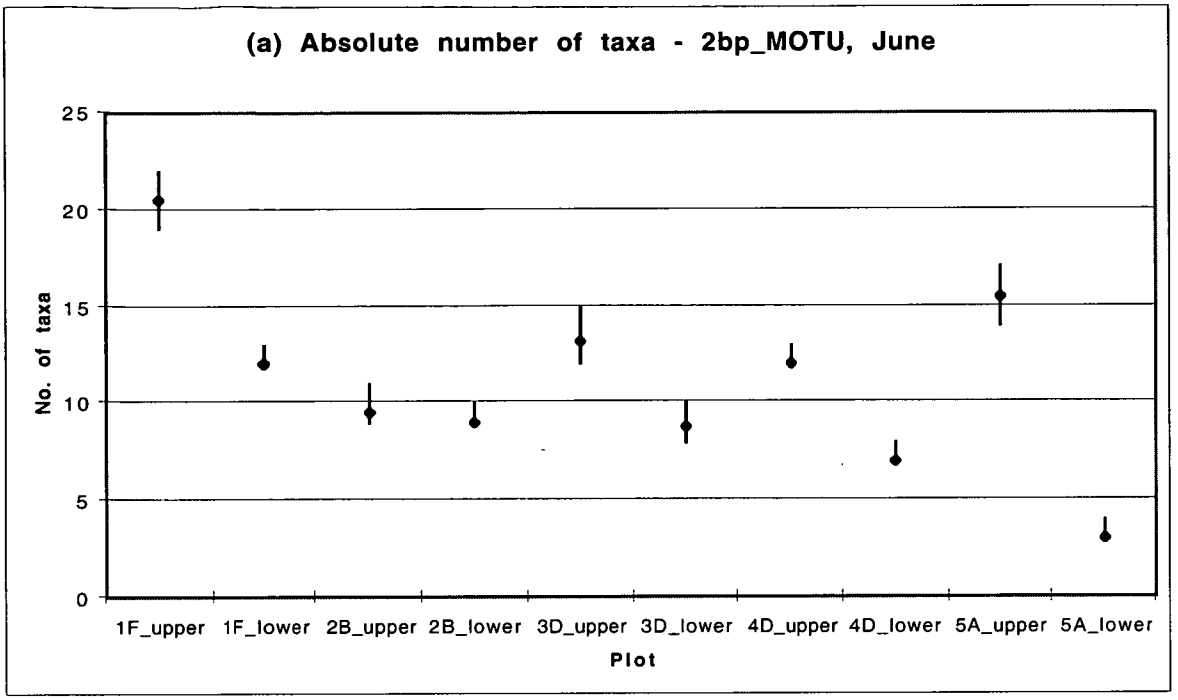
As well as the three diversity indices described above, three other parameters of interest were calculated from the MOTU dataset. The first and most basic was the absolute number of taxa per sample. Also calculated was the number of *unique* taxa per sample, i.e. MOTU which are found only this sample and no other. Tautologically, any taxon with only a single member is unique to the sample in which it was found, therefore counting these taxa as unique to a certain plot might seem a relatively uninteresting piece of information (though it may be informative to know if certain plots contain significantly more singletons than others). If, on the other hand, a taxon has multiple members and is still confined to a single plot then this enables us to make some statement about its distribution. Therefore another parameter given for each plot is the number of unique taxa with more than one member.

### 6.2.2 Estimation of diversity indices from Sourhope MOTU data

As discussed in the previous chapter, 100 runs were carried out on the entire Sourhope survey dataset, with 2bp variation or less allowed between any two sequences within a MOTU, and from each run a list of sequences with their MOTU assignments was produced. Using the perl script **div_table.pl** the sequences were divided into their plot of origin and sampling date, so that for each sample, and for each of the 100 runs, a number of individuals and a set of taxon assignments was obtained, from which the various diversity indices could be calculated. The script **div_indices.pl** was used to derive, for each sample, six parameters: number of taxa, number of unique taxa, number of unique non-singleton taxa, Shannon index, Simpson index, and dominance value. For each of these the minimum, maximum and mean value from all 100 runs was determined. A summary of results is shown in Table 6.2.1, and graphs plotted in Figures 6.2.1 and 6.2.2. The number of individuals sampled per plot is also included. The June samples, which were analysed in upper and lower horizons, are included both combined and split into horizons.

| | | | | Number of taxa | | | Unique taxa | | | Unique non-singletons | | | Shannon Index | | | Simpson Index | | | Dominance Index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample | Date | Treatment | No. Indiv. | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean |
| 1F | June | Control | 101 | 25 | 28 | 26.48 | 7 | 10 | 7.65 | 0 | 1 | 0.17 | 2.409 | 2.508 | 2.444 | 1.999 | 2.059 | 2.009 | 0.267 | 0.277 | 0.277 |
| 2B | June | Control | 97 | 12 | 14 | 12.57 | 3 | 4 | 3.05 | 0 | 0 | 0 | 1.776 | 1.903 | 1.798 | 1.458 | 1.521 | 1.464 | 0.402 | 0.402 | 0.402 |
| 3D | June | Control | 118 | 14 | 17 | 15.38 | 2 | 6 | 3.12 | 0 | 0 | 0 | 1.825 | 1.919 | 1.867 | 1.404 | 1.444 | 1.412 | 0.441 | 0.449 | 0.449 |
| 4D | June | Control | 123 | 13 | 15 | 13.87 | 1 | 2 | 1.02 | 0 | 0 | 0 | 0.932 | 0.989 | 0.946 | 0.431 | 0.451 | 0.433 | 0.797 | 0.805 | 0.804 |
| 5A | June | Control | 66 | 14 | 18 | 15.52 | 1 | 3 | 1.14 | 0 | 0 | 0 | 2.147 | 2.319 | 2.213 | 1.944 | 2.036 | 1.968 | 0.242 | 0.258 | 0.257 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1F_upper | June | Control | 51 | 19 | 22 | 20.46 | 6 | 8 | 6.64 | 0 | 1 | 0.17 | 2.516 | 2.654 | 2.580 | 2.380 | 2.478 | 2.419 | 0.196 | 0.196 | 0.196 |
| 1F_lower | June | Control | 50 | 12 | 13 | 12.02 | 1 | 2 | 1.01 | 0 | 0 | 0 | 1.816 | 1.894 | 1.818 | 1.512 | 1.577 | 1.513 | 0.340 | 0.360 | 0.360 |
| 2B_upper | June | Control | 47 | 9 | 11 | 9.55 | 1 | 2 | 1.03 | 0 | 0 | 0 | 1.598 | 1.814 | 1.641 | 1.365 | 1.493 | 1.381 | 0.426 | 0.426 | 0.426 |
| 2B_lower | June | Control | 50 | 9 | 10 | 9.02 | 2 | 3 | 2.02 | 0 | 0 | 0 | 1.725 | 1.779 | 1.726 | 1.542 | 1.562 | 1.543 | 0.380 | 0.380 | 0.380 |
| 3D_upper | June | Control | 73 | 12 | 15 | 13.22 | 2 | 4 | 2.6 | 0 | 0 | 0 | 1.790 | 1.892 | 1.827 | 1.454 | 1.475 | 1.461 | 0.411 | 0.411 | 0.411 |
| 3D_lower | June | Control | 45 | 8 | 10 | 8.82 | 0 | 2 | 0.52 | 0 | 0 | 0 | 1.515 | 1.649 | 1.551 | 1.207 | 1.292 | 1.215 | 0.489 | 0.511 | 0.511 |
| 4D_upper | June | Control | 52 | 12 | 13 | 12.07 | 1 | 1 | 1 | 0 | 0 | 0 | 1.255 | 1.344 | 1.261 | 0.680 | 0.735 | 0.684 | 0.692 | 0.712 | 0.710 |
| 4D_lower | June | Control | 71 | 7 | 8 | 7.05 | 0 | 1 | 0.02 | 0 | 0 | 0 | 0.593 | 0.665 | 0.598 | 0.271 | 0.304 | 0.273 | 0.859 | 0.873 | 0.873 |
| 5A_upper | June | Control | 44 | 14 | 17 | 15.5 | 1 | 3 | 1.13 | 0 | 0 | 0 | 2.277 | 2.473 | 2.374 | 2.170 | 2.330 | 2.237 | 0.250 | 0.250 | 0.250 |
| 5A_lower | June | Control | 22 | 3 | 4 | 3.02 | 0 | 1 | 0.01 | 0 | 0 | 0 | 0.937 | 1.097 | 0.939 | 0.878 | 1.012 | 0.880 | 0.545 | 0.591 | 0.590 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1E | Oct | Biocide | 54 | 9 | 9 | 9 | 2 | 2 | 2 | 0 | 0 | 0 | 0.791 | 0.791 | 0.791 | 0.367 | 0.367 | 0.367 | 0.833 | 0.833 | 0.833 |
| 2D | Oct | Biocide | 36 | 9 | 11 | 9.75 | 0 | 1 | 0.04 | 0 | 0 | 0 | 1.349 | 1.450 | 1.378 | 0.881 | 0.897 | 0.884 | 0.639 | 0.639 | 0.639 |
| 3C | Oct | Biocide | 52 | 13 | 16 | 14.52 | 2 | 4 | 2.57 | 0 | 0 | 0 | 1.915 | 2.270 | 2.044 | 1.541 | 1.964 | 1.641 | 0.327 | 0.423 | 0.405 |
| 4B | Oct | Biocide | 57 | 20 | 26 | 23.01 | 13 | 18 | 15.25 | 0 | 3 | 1.94 | 2.162 | 2.771 | 2.534 | 1.456 | 2.539 | 2.088 | 0.193 | 0.474 | 0.311 |
| 5E | Oct | Biocide | 63 | 13 | 16 | 14.17 | 0 | 1 | 0.04 | 0 | 0 | 0 | 1.906 | 2.050 | 1.969 | 1.601 | 1.642 | 1.620 | 0.365 | 0.365 | 0.365 |
| 1F | Oct | Control | 79 | 16 | 19 | 17.36 | 4 | 5 | 4.11 | 1 | 1 | 1 | 1.960 | 2.122 | 2.029 | 1.565 | 1.648 | 1.592 | 0.354 | 0.367 | 0.365 |
| 2B | Oct | Control | 57 | 13 | 16 | 14.37 | 2 | 3 | 2.05 | 0 | 0 | 0 | 1.768 | 1.888 | 1.806 | 1.425 | 1.486 | 1.434 | 0.368 | 0.386 | 0.384 |
| 3D | Oct | Control | 66 | 11 | 14 | 12.18 | 2 | 2 | 2 | 0 | 0 | 0 | 1.599 | 1.700 | 1.644 | 1.243 | 1.257 | 1.250 | 0.470 | 0.470 | 0.470 |
| 4D | Oct | Control | 45 | 7 | 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.946 | 0.946 | 0.946 | 0.552 | 0.552 | 0.552 | 0.756 | 0.756 | 0.756 |
| 1B | Oct | Lime | 69 | 9 | 10 | 9.03 | 0 | 1 | 0.03 | 0 | 0 | 0 | 1.170 | 1.190 | 1.170 | 0.729 | 0.730 | 0.729 | 0.681 | 0.681 | 0.681 |
| 2C | Oct | Lime | 46 | 8 | 11 | 8.76 | 1 | 4 | 1.75 | 0 | 0 | 0 | 1.446 | 1.717 | 1.516 | 1.110 | 1.336 | 1.166 | 0.478 | 0.543 | 0.527 |
| 3B | Oct | Lime | 66 | 11 | 13 | 11.8 | 1 | 2 | 1.1 | 0 | 0 | 0 | 1.605 | 1.658 | 1.633 | 1.179 | 1.197 | 1.195 | 0.515 | 0.515 | 0.515 |
| 4F | Oct | Lime | 58 | 15 | 17 | 15.23 | 4 | 6 | 4.78 | 0 | 1 | 0.15 | 1.572 | 1.684 | 1.584 | 0.896 | 0.954 | 0.902 | 0.621 | 0.638 | 0.636 |
| 5B | Oct | Lime | 79 | 14 | 15 | 14.99 | 2 | 2 | 2 | 1 | 1 | 1 | 1.883 | 2.044 | 2.042 | 1.511 | 1.679 | 1.677 | 0.329 | 0.329 | 0.329 |
| 1C | Oct | Nitrogen | 64 | 13 | 15 | 13.82 | 0 | 0 | 0 | 0 | 0 | 0 | 1.956 | 2.062 | 1.993 | 1.671 | 1.737 | 1.685 | 0.297 | 0.297 | 0.297 |
| 2A | Oct | Nitrogen | 73 | 14 | 16 | 14.16 | 2 | 3 | 2.08 | 0 | 0 | 0 | 1.936 | 2.000 | 1.942 | 1.642 | 1.692 | 1.644 | 0.342 | 0.356 | 0.356 |
| 3F | Oct | Nitrogen | 63 | 14 | 17 | 15.25 | 1 | 4 | 2.38 | 0 | 0 | 0 | 1.751 | 1.939 | 1.818 | 1.144 | 1.265 | 1.174 | 0.524 | 0.556 | 0.549 |
| 4E | Oct | Nitrogen | 75 | 14 | 18 | 15.7 | 4 | 7 | 4.78 | 0 | 1 | 0.15 | 1.583 | 1.889 | 1.663 | 1.039 | 1.374 | 1.085 | 0.467 | 0.573 | 0.560 |
| 5C | Oct | Nitrogen | 24 | 8 | 9 | 8.02 | 0 | 1 | 0.2 | 0 | 0 | 0 | 1.965 | 2.059 | 1.967 | 2.186 | 2.288 | 2.188 | 0.167 | 0.167 | 0.167 |
| 1A | Oct | Nitrogen + lime | 68 | 25 | 28 | 25.5 | 7 | 10 | 7.58 | 0 | 1 | 0.05 | 2.697 | 2.899 | 2.726 | 2.259 | 2.662 | 2.300 | 0.221 | 0.294 | 0.287 |
| 2E | Oct | Nitrogen + lime | 81 | 14 | 14 | 14 | 4 | 4 | 4 | 0 | 0 | 0 | 1.812 | 1.812 | 1.812 | 1.455 | 1.455 | 1.455 | 0.395 | 0.395 | 0.395 |
| 3A | Oct | Nitrogen + lime | 66 | 20 | 22 | 20.9 | 1 | 4 | 2.42 | 0 | 2 | 0.75 | 2.541 | 2.724 | 2.631 | 2.333 | 2.559 | 2.452 | 0.212 | 0.212 | 0.212 |
| 4C | Oct | Nitrogen + lime | 66 | 14 | 17 | 14.97 | 2 | 4 | 2.4 | 0 | 0 | 0 | 1.892 | 2.065 | 1.956 | 1.525 | 1.652 | 1.563 | 0.364 | 0.394 | 0.387 |
| 5D | Oct | Nitrogen + lime | 52 | 10 | 13 | 11.5 | 1 | 3 | 1.6 | 0 | 0 | 0 | 1.620 | 1.780 | 1.692 | 1.290 | 1.370 | 1.323 | 0.442 | 0.462 | 0.454 |

Table 6.2.1 Diversity data for 2bp_MOTU from all June and October plots, showing mean, maximum and minimum values from 100 runs of **define_MOTU.pl**



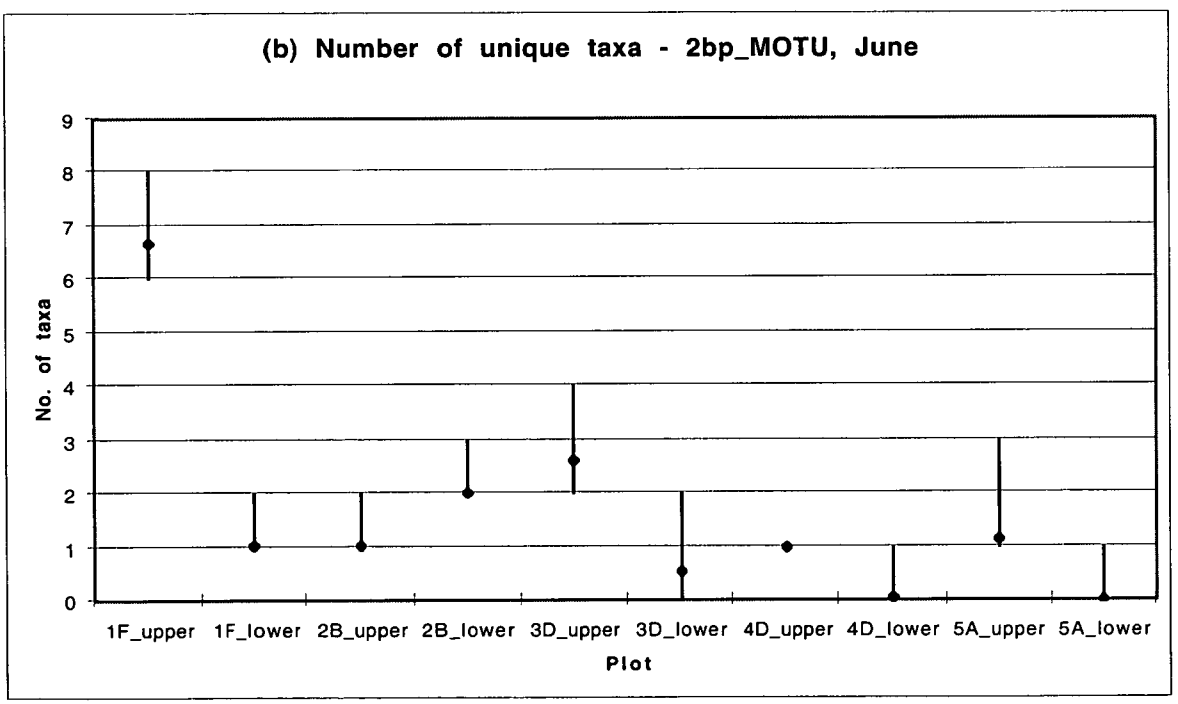
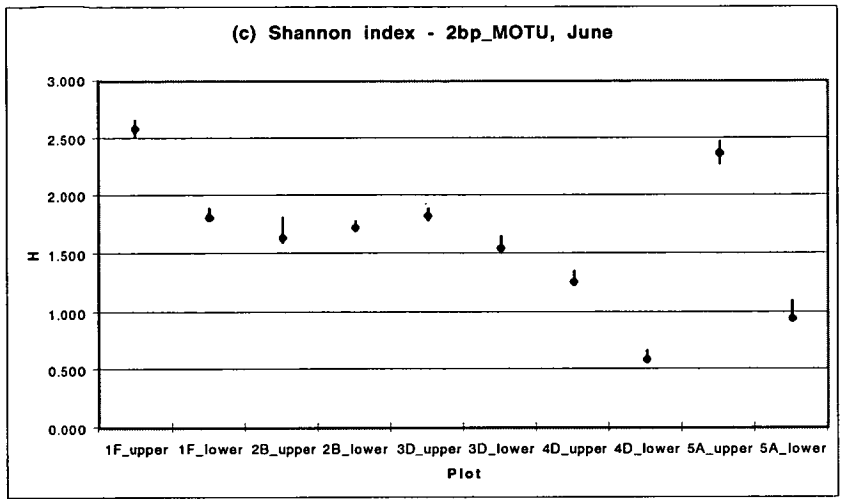
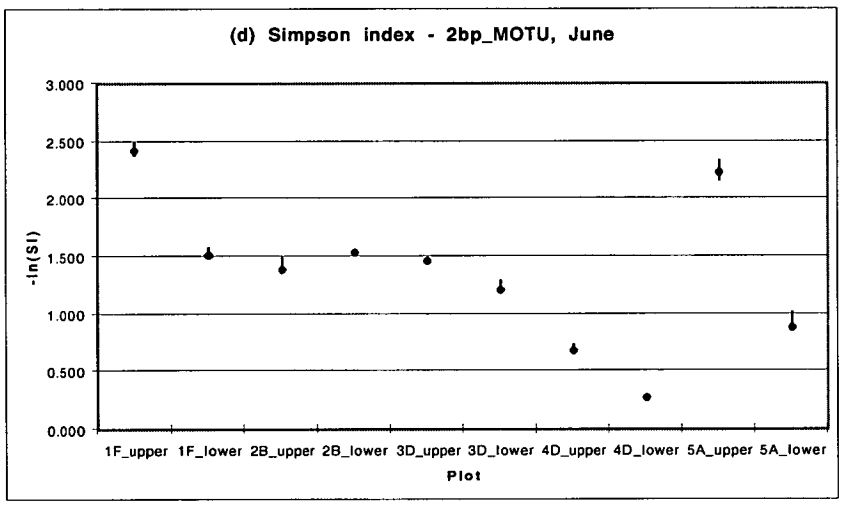




Figure 6.2.1 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the June 2001 samples for MOTU with 2bp variation, showing mean and range over 100 runs.



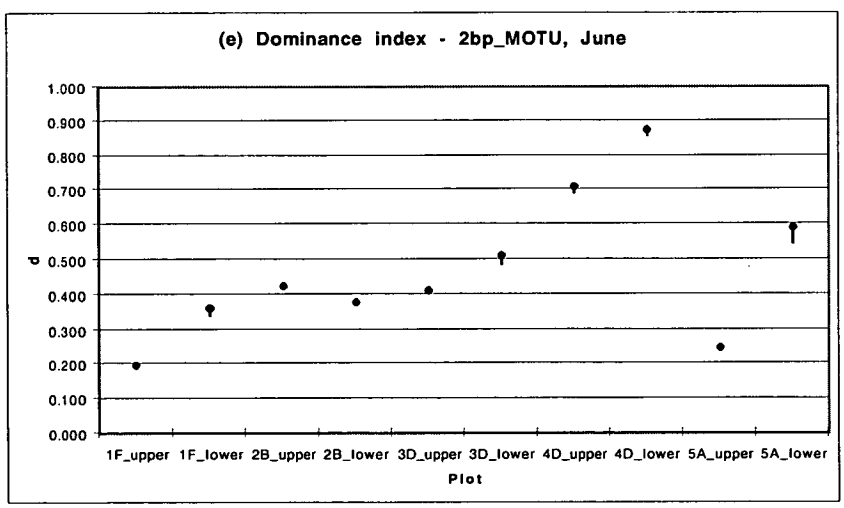
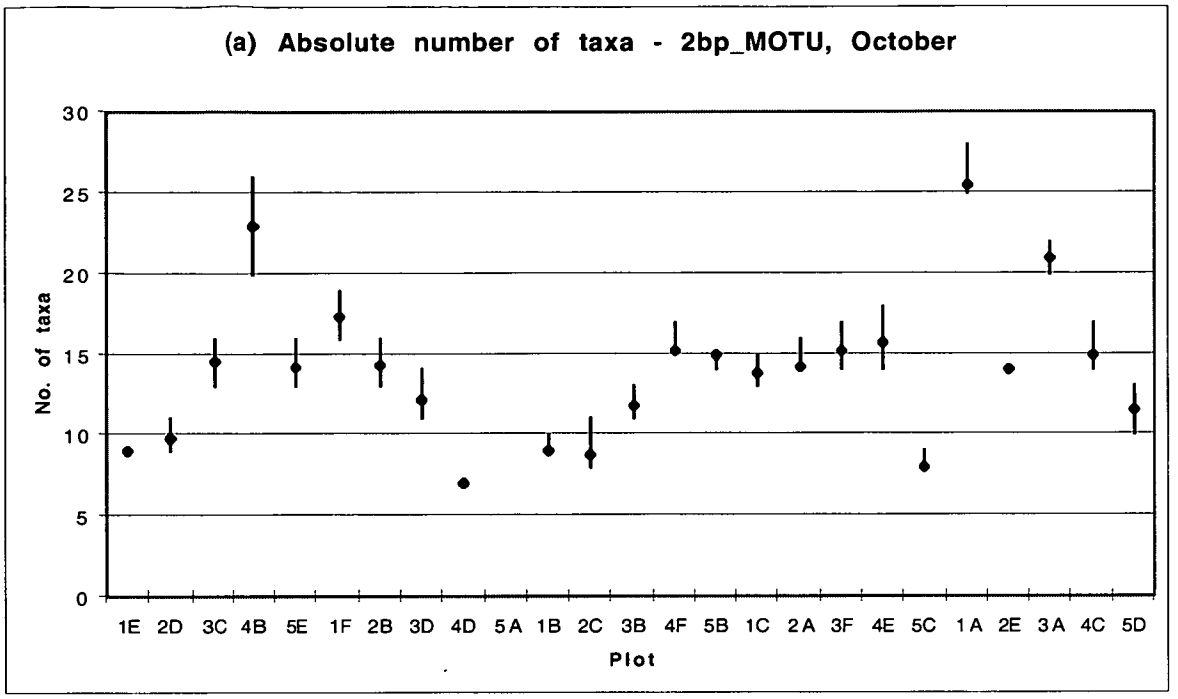




Figure 6.2.1 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the June 2001 samples for MOTU with 2bp variation, showing mean and range over 100 runs.



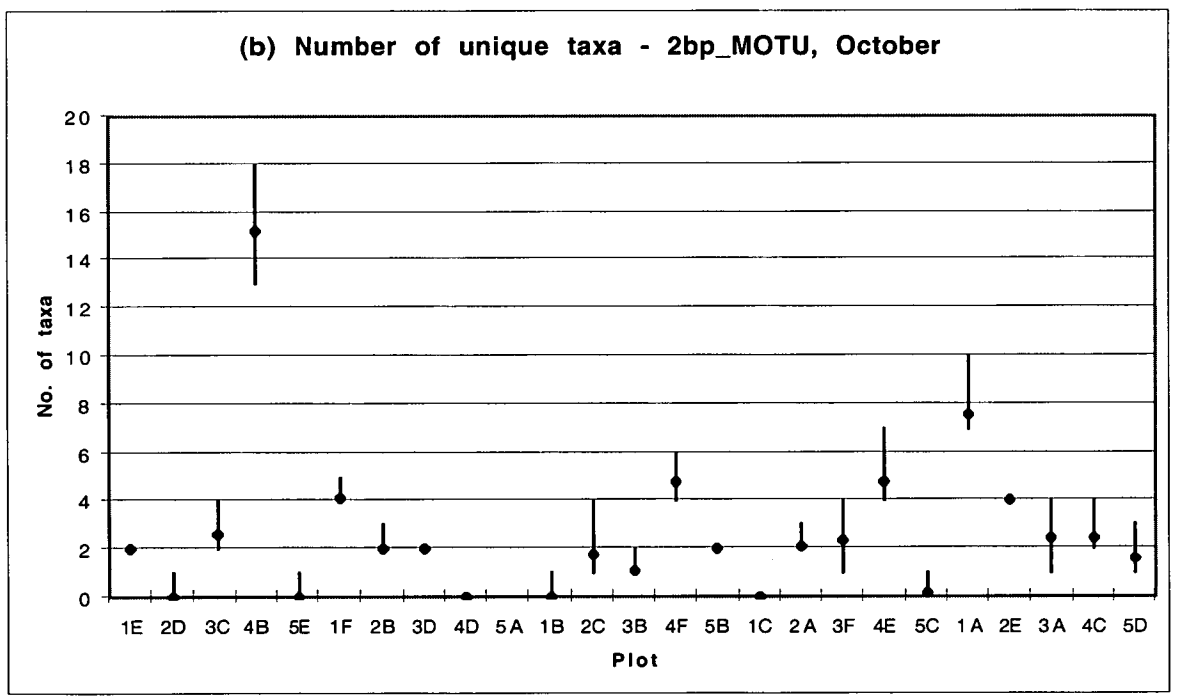
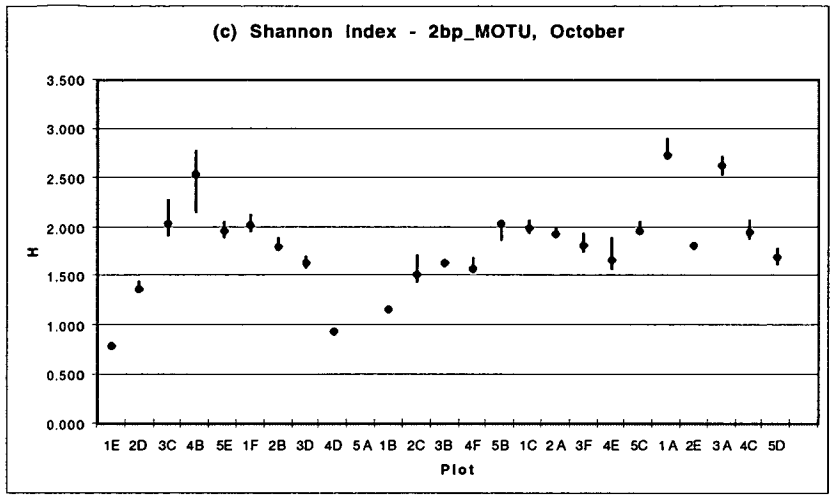
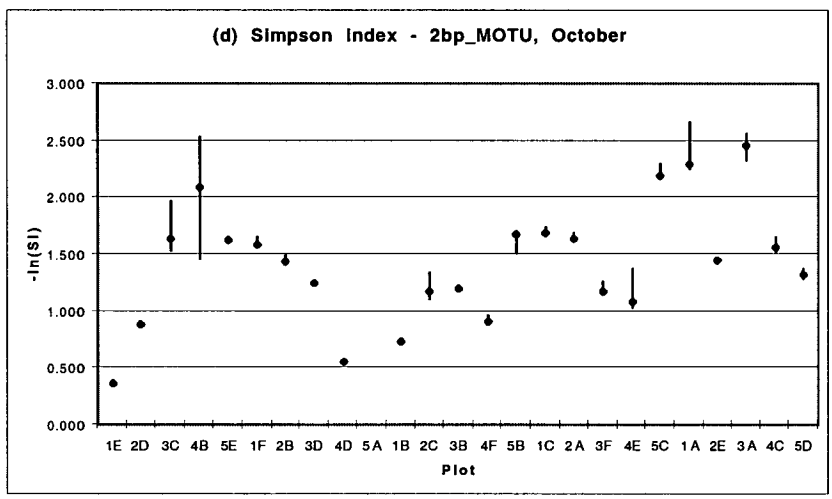




Figure 6.2.2 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the October 2001 samples for MOTU with 2bp variation, showing mean and range over 100 runs.



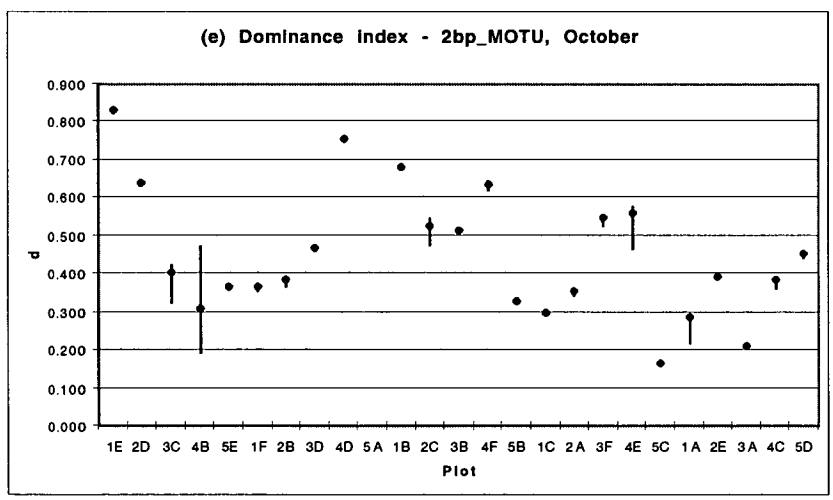




Figure 6.2.2 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the October 2001 samples for MOTU with 2bp variation, showing mean and range over 100 runs.

The three diversity indices show some general patterns. For the June samples, all of the indices agree that 1F_upper shows the highest diversity and 4D_lower the least, and also that the upper horizon is always more diverse than the corresponding lower horizon, with the exception of 2B where the reverse is seen (however these values are similar enough that they may not be significantly different). The rank orders given by the three indices for the June samples are as follows (most diverse first; u=upper horizon, l=lower horizon):

Shannon - 1Fu,5Au,3Du,1Fl,2Bl,2Bu,3Dl,4Du,5Al,4Dl
Simpson - 1Fu,5Au,2Bl,1Fl,3Du,2Bu,3Dl,5Al,4Du,4Dl
Dominance - 1Fu,5Au,1Fl,2Bl,3Du,2Bu,3Dl,5Al,4Du,4Dl

In the October set, in which all plots were sampled but none were split by horizon, the indices disagree about which plot is the most diverse (1A by the Shannon index, 3A by Simpson, and 5C by dominance) but all agree that 1E is the least diverse. The rank orders are:

Shannon -
1A,3A,4B,3C,5B,1F,1C,5E,5C,4C,2A,3F,2E,2B,5D,4E,3D,3B,4F,2C,2D,1B,4D,1E
Simpson -
3A,1A,5C,4B,1C,5B,2A,3C,5E,1F,4C,2E,2B,5D,3D,3B,3F,2C,4E,4F,2D,1B,4D,1E
Dominance -
5C,3A,1A,1C,4B,5B,2A,5E,1F,2B,4C,2E,3C,5D,3D,3B,2C,3F,4E,4F,2D,1B,4D,1E

In the graphs in Figures 6.2.1 and 6.2.2, the range of values resulting from the 100 runs are shown as error bars above and below each point, representing the minimum and maximum value observed. These ranges are small in most cases, but in a few instances are noticeably larger, particularly plots 3C, 4E, 4B and 1A in October, and here the variation is most strongly reflected in the Simpson index values. In the majority of cases, however, it is notable that the error bars on the number-of-taxa graphs are proportionately larger than the corresponding error bars on the index values derived from these numbers. In most cases, therefore, variation in numbers of taxa resulting from sequence-processing-order artefacts in the MOTU clustering process are much more weakly reflected in the resulting diversity index values. This is due to the fact that, as was found in Chapter 5, most of the variation in taxon number is caused by differences in the number of singleton taxa, and singletons do not strongly influence any of these diversity indices. In the Simpson index, for example, the top line of the equation, "$n^2$-$n$", will always equal zero for taxa containing one member ($n$=1). Therefore variations in the number of singletons should not affect the value of this index at all, except in so far as they consequently alter the membership of the more abundant taxa by moving individuals from these taxa to singletons and vice versa; this should normally be a weak effect. The same is true of the dominance index, whose value depends *only* on the membership of the single

most abundant taxon. The exceptional cases, then, where large fluctuations in these indices are seen, indicate that something more unusual is occurring.

An examination of the MOTU in plot 4B indicates that the dominant taxon fluctuates between containing as few as 11 and as many as 27 members. This indicates that there is a group of closely related sequences all originating within this plot, which are split or lumped depending on searching order. The sequences in question all belong to the *Helicotylenchus* group discussed in the previous chapter. 16515ED is an example of a divergent sequence, differing from the dominant type by 5 bases; if this sequence is tested early on in the process, it defines a separate MOTU which subsequently 'pulls' a significant number of intermediate sequences in with it, while if it is processed late it either remains a singleton, or may itself be 'pulled' into a more common MOTU by an intermediate sequence. There is another MOTU, repeatedly found, containing five sequences (16507ED, 16516ED, 16546ED, 16574ED and 16582ED), all showing identical variations at the three significantly variable sites 172, 270 and 413 – and which is found in no other sample in the entire dataset, i.e. this MOTU is unique to plot 4B. Here we see sequence variation which is certainly not due to random errors but to a distinct taxon with a localised distribution – yet, because it differs from the dominant sequence type by only 3 bases, it is only sometimes recognised as a distinct taxon at all. This plot appears to contain a number of distinct *Helicotylenchus*-like sequences found nowhere else, perhaps a genuine 'species flock', causing this fluctuating behaviour in its diversity index values. This may be seen as an illustration of how the variability of taxon assignment, which might be considered a disadvantage of the MOTU approach, can nonetheless serve as a means of drawing our attention to potentially interesting biological phenomena which might otherwise escape our notice.

Plot 4B is also found to contain far more unique taxa than any other (see Figure 6.2.2 (b)), with a mean of approximately 15 (out of only ~23 taxa in total) and a range from 13 to 18 in different runs; the next highest, 1A, only contains around 8 unique taxa (mean 7.58, range 7-10). Up to 3 of the unique taxa in plot 4B are found to contain more than one member, and this plot also contains substantially more singletons than any other. This further information suggests that the composition of the nematode community in this plot is unusual and distinct from other sites within the field, for reasons which are not yet clear.

There are other MOTU which are both unique to certain plots and contain more than one member, in which sequences are found to cluster together repeatedly in different runs. From plot 1F, a two-member MOTU was found containing sequences 15408ED and 15483ED (which belong to 2bp_MOTU0097 in the main tree shown in figure 4.6, within the dorylaimid group); from plot 3A, a five-member MOTU containing 16013ED, 16015ED, 16067ED, 16071ED and 16073ED (MOTU0051 on the tree, also a dorylaimid taxon); from plot 4B, the 8-member *Helicotylenchus*-like MOTU already mentioned, 16507ED, 16515ED, 16516ED, 16528ED, 16546ED, 16556ED, 16574ED, and 16582ED (2bp_MOTU0046 on the tree), and an additional two-member MOTU also similar to *Helicotylenchus* comprising 16524ED and 16558ED (2bp_MOTU0038 on the tree); and a three member MOTU found in plot 5B, comprising 17132ED, 17136ED and 17190ED (2bp_MOTU0063 on the tree, a mononchid close to *Mylonchulus*

*arenicolus*). These results suggest that the distribution of these taxa may be driven by features peculiar to these particular plots, such as the presence of a particular plant or fungal species.

## 6.3 Effect of varying MOTU designation threshold

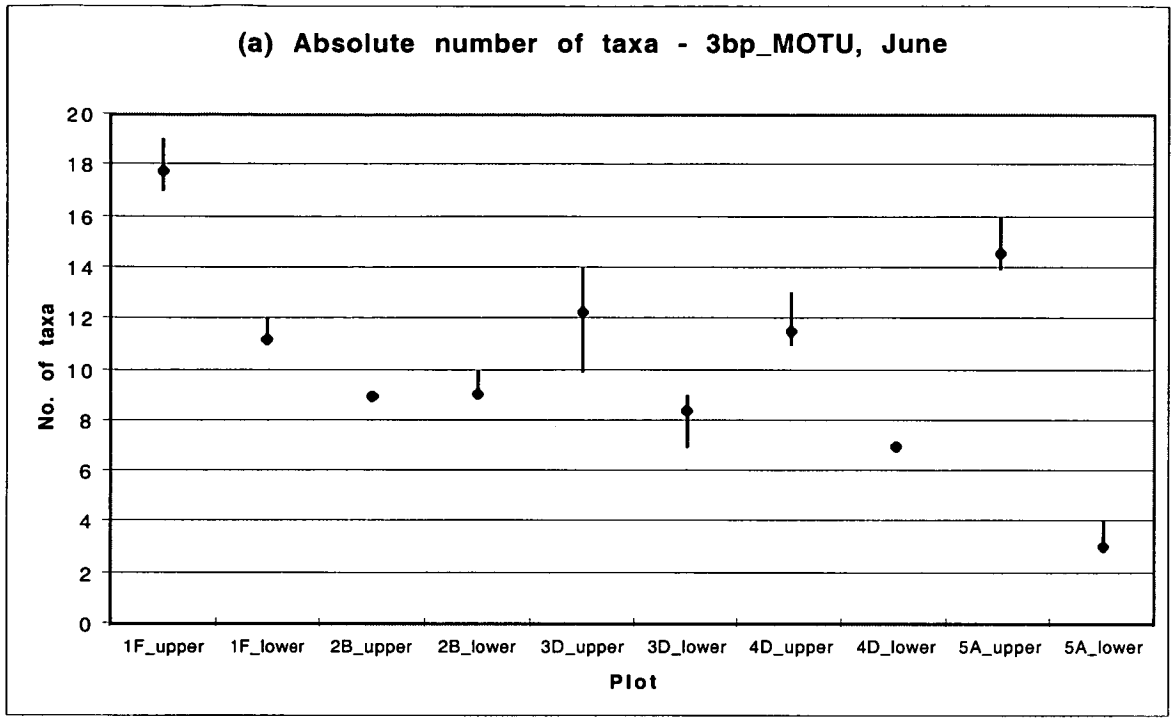
The dataset was also reanalysed 100 times with 3 and 4 bases as the level of variation allowed within a MOTU; the same set of parameters (number of taxa, unique taxa, unique non-singletons, Shannon index, Simpson index, dominance value) were derived from each set. Results are shown in Tables 6.3.1-6.3.2, and plotted in Figures 6.3.1-6.3.4.

| Sample | Date | Treatment | No. Indiv. | Number of taxa | | | Unique taxa | | | Unique non-singletons | | | Shannon Index | | | Simpson Index | | | Dominance Index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean |
| 1F | June | Control | 101 | 22 | 25 | 22.97 | 3 | 6 | 4.63 | 0 | 1 | 0.15 | 2.312 | 2.425 | 2.348 | 1.931 | 2.005 | 1.953 | 0.277 | 0.287 | 0.284 |
| 2B | June | Control | 97 | 12 | 13 | 12.03 | 3 | 4 | 3.03 | 0 | 0 | 0 | 1.776 | 1.819 | 1.777 | 1.458 | 1.479 | 1.459 | 0.402 | 0.402 | 0.402 |
| 3D | June | Control | 118 | 12 | 16 | 14.3 | 2 | 5 | 2.69 | 0 | 0 | 0 | 1.749 | 1.885 | 1.835 | 1.388 | 1.414 | 1.408 | 0.449 | 0.449 | 0.449 |
| 4D | June | Control | 123 | 12 | 15 | 12.86 | 0 | 1 | 0.47 | 0 | 0 | 0 | 0.886 | 0.989 | 0.913 | 0.411 | 0.451 | 0.421 | 0.797 | 0.813 | 0.809 |
| 5A | June | Control | 66 | 14 | 16 | 14.55 | 1 | 2 | 1.1 | 0 | 0 | 0 | 2.128 | 2.227 | 2.172 | 1.931 | 1.970 | 1.953 | 0.258 | 0.258 | 0.258 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1F_upper | June | Control | 51 | 17 | 19 | 17.77 | 3 | 6 | 4.43 | 0 | 1 | 0.15 | 2.423 | 2.543 | 2.470 | 2.291 | 2.406 | 2.333 | 0.196 | 0.216 | 0.209 |
| 1F_lower | June | Control | 50 | 11 | 12 | 11.2 | 0 | 1 | 0.2 | 0 | 0 | 0 | 1.742 | 1.816 | 1.756 | 1.458 | 1.512 | 1.469 | 0.360 | 0.360 | 0.360 |
| 2B_upper | June | Control | 47 | 9 | 9 | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 1.598 | 1.598 | 1.598 | 1.365 | 1.365 | 1.365 | 0.426 | 0.426 | 0.426 |
| 2B_lower | June | Control | 50 | 9 | 10 | 9.03 | 2 | 3 | 2.03 | 0 | 0 | 0 | 1.725 | 1.792 | 1.726 | 1.542 | 1.581 | 1.543 | 0.380 | 0.380 | 0.380 |
| 3D_upper | June | Control | 73 | 10 | 14 | 12.29 | 2 | 3 | 2.2 | 0 | 0 | 0 | 1.707 | 1.853 | 1.797 | 1.436 | 1.465 | 1.455 | 0.411 | 0.411 | 0.411 |
| 3D_lower | June | Control | 45 | 7 | 9 | 8.37 | 0 | 2 | 0.49 | 0 | 0 | 0 | 1.465 | 1.557 | 1.530 | 1.197 | 1.214 | 1.210 | 0.511 | 0.511 | 0.511 |
| 4D_upper | June | Control | 52 | 11 | 13 | 11.5 | 0 | 1 | 0.47 | 0 | 0 | 0 | 1.166 | 1.344 | 1.211 | 0.626 | 0.735 | 0.653 | 0.692 | 0.731 | 0.721 |
| 4D_lower | June | Control | 71 | 7 | 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.593 | 0.593 | 0.593 | 0.271 | 0.271 | 0.271 | 0.873 | 0.873 | 0.873 |
| 5A_upper | June | Control | 44 | 14 | 16 | 14.54 | 1 | 2 | 1.09 | 0 | 0 | 0 | 2.247 | 2.397 | 2.313 | 2.134 | 2.247 | 2.195 | 0.250 | 0.250 | 0.250 |
| 5A_lower | June | Control | 22 | 3 | 4 | 3.01 | 0 | 1 | 0.01 | 0 | 0 | 0 | 0.937 | 1.024 | 0.938 | 0.878 | 0.899 | 0.878 | 0.591 | 0.591 | 0.591 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1E | Oct | Biocide | 54 | 9 | 9 | 9 | 2 | 2 | 2 | 0 | 0 | 0 | 0.791 | 0.791 | 0.791 | 0.367 | 0.367 | 0.367 | 0.833 | 0.833 | 0.833 |
| 2D | Oct | Biocide | 36 | 8 | 10 | 9.19 | 0 | 0 | 0 | 0 | 0 | 0 | 1.296 | 1.388 | 1.355 | 0.874 | 0.885 | 0.882 | 0.639 | 0.639 | 0.639 |
| 3C | Oct | Biocide | 52 | 11 | 14 | 13.29 | 0 | 1 | 0.1 | 0 | 0 | 0 | 1.673 | 1.995 | 1.939 | 1.364 | 1.573 | 1.550 | 0.423 | 0.423 | 0.423 |
| 4B | Oct | Biocide | 57 | 16 | 22 | 18.71 | 8 | 13 | 10.3 | 0 | 2 | 0.71 | 1.861 | 2.541 | 2.152 | 1.242 | 2.182 | 1.592 | 0.246 | 0.526 | 0.427 |
| 5E | Oct | Biocide | 63 | 12 | 15 | 13.33 | 0 | 0 | 0 | 0 | 0 | 0 | 1.785 | 2.010 | 1.917 | 1.527 | 1.632 | 1.601 | 0.365 | 0.365 | 0.365 |
| 1F | Oct | Control | 79 | 14 | 17 | 14.97 | 2 | 3 | 2.42 | 1 | 1 | 1 | 1.854 | 2.016 | 1.904 | 1.492 | 1.570 | 1.519 | 0.367 | 0.380 | 0.374 |
| 2B | Oct | Control | 57 | 11 | 13 | 11.48 | 0 | 1 | 0.06 | 0 | 0 | 0 | 1.623 | 1.720 | 1.637 | 1.314 | 1.371 | 1.318 | 0.404 | 0.421 | 0.420 |
| 3D | Oct | Control | 66 | 9 | 13 | 11.34 | 1 | 1 | 1 | 0 | 0 | 0 | 1.426 | 1.679 | 1.609 | 1.183 | 1.256 | 1.244 | 0.470 | 0.470 | 0.470 |
| 4D | Oct | Control | 45 | 7 | 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.946 | 0.946 | 0.946 | 0.552 | 0.552 | 0.552 | 0.756 | 0.756 | 0.756 |
| 1B | Oct | Lime | 69 | 8 | 10 | 8.92 | 0 | 1 | 0.02 | 0 | 0 | 0 | 1.121 | 1.190 | 1.165 | 0.723 | 0.730 | 0.728 | 0.681 | 0.681 | 0.681 |
| 2C | Oct | Lime | 46 | 8 | 9 | 8.11 | 1 | 2 | 1.06 | 0 | 0 | 0 | 1.446 | 1.538 | 1.456 | 1.110 | 1.183 | 1.118 | 0.522 | 0.543 | 0.541 |
| 3B | Oct | Lime | 66 | 9 | 12 | 11.19 | 1 | 1 | 1 | 0 | 0 | 0 | 1.438 | 1.637 | 1.610 | 1.117 | 1.195 | 1.189 | 0.515 | 0.515 | 0.515 |
| 4F | Oct | Lime | 58 | 13 | 15 | 13.64 | 2 | 5 | 3.15 | 0 | 0 | 0 | 1.453 | 1.572 | 1.483 | 0.838 | 0.896 | 0.847 | 0.638 | 0.655 | 0.653 |
| 5B | Oct | Lime | 79 | 12 | 14 | 13.85 | 1 | 1 | 1 | 1 | 1 | 1 | 1.774 | 2.026 | 2.003 | 1.436 | 1.677 | 1.653 | 0.329 | 0.354 | 0.329 |
| 1C | Oct | Nitrogen | 64 | 13 | 14 | 13.42 | 0 | 0 | 0 | 0 | 0 | 0 | 1.956 | 2.000 | 1.974 | 1.671 | 1.687 | 1.678 | 0.297 | 0.297 | 0.297 |
| 2A | Oct | Nitrogen | 73 | 12 | 15 | 13.38 | 1 | 2 | 1.32 | 0 | 0 | 0 | 1.811 | 1.955 | 1.919 | 1.541 | 1.655 | 1.638 | 0.356 | 0.384 | 0.356 |
| 3F | Oct | Nitrogen | 63 | 12 | 15 | 13.64 | 1 | 2 | 1.2 | 0 | 0 | 0 | 1.639 | 1.795 | 1.713 | 1.083 | 1.151 | 1.104 | 0.556 | 0.571 | 0.568 |
| 4E | Oct | Nitrogen | 75 | 14 | 17 | 14.6 | 3 | 5 | 3.36 | 0 | 1 | 0.05 | 1.583 | 1.884 | 1.607 | 1.039 | 1.377 | 1.050 | 0.467 | 0.573 | 0.570 |
| 5C | Oct | Nitrogen | 24 | 8 | 9 | 8.02 | 0 | 1 | 0.05 | 0 | 0 | 0 | 1.965 | 2.059 | 1.967 | 2.186 | 2.288 | 2.188 | 0.167 | 0.167 | 0.167 |
| 1A | Oct | Nitrogen + lime | 68 | 24 | 26 | 24.19 | 6 | 8 | 6.26 | 0 | 1 | 0.01 | 2.655 | 2.802 | 2.666 | 2.234 | 2.538 | 2.247 | 0.235 | 0.294 | 0.292 |
| 2E | Oct | Nitrogen + lime | 81 | 11 | 14 | 12.89 | 3 | 4 | 3.01 | 0 | 0 | 0 | 1.516 | 1.812 | 1.775 | 1.220 | 1.455 | 1.447 | 0.395 | 0.395 | 0.395 |
| 3A | Oct | Nitrogen + lime | 66 | 17 | 21 | 19.18 | 0 | 2 | 0.78 | 0 | 1 | 0.06 | 2.389 | 2.612 | 2.501 | 2.228 | 2.456 | 2.299 | 0.212 | 0.212 | 0.212 |
| 4C | Oct | Nitrogen + lime | 66 | 12 | 17 | 14.24 | 1 | 4 | 1.27 | 0 | 0 | 0 | 1.722 | 2.065 | 1.922 | 1.435 | 1.660 | 1.545 | 0.364 | 0.394 | 0.391 |
| 5D | Oct | Nitrogen + lime | 52 | 10 | 13 | 10.54 | 1 | 3 | 1.2 | 0 | 0 | 0 | 1.620 | 1.770 | 1.645 | 1.290 | 1.367 | 1.300 | 0.442 | 0.462 | 0.459 |

Table 6.3.1 Diversity data for 3bp_MOTU from all June and October plots, showing mean, maximum and minimum values from 100 runs of **define_MOTU.pl**

| Sample | Date | Treatment | No. indiv. | Number of taxa | | | Unique taxa | | | Unique non-singletons | | | Shannon Index | | | Simpson Index | | | Dominance Index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean |
| 1F_June | June | Control | 101 | 21 | 22 | 21.08 | 2 | 4 | 2.73 | 0 | 0 | 0 | 2.290 | 2.333 | 2.292 | 1.927 | 1.966 | 1.928 | 0.277 | 0.287 | 0.287 |
| 2B_June | June | Control | 97 | 11 | 13 | 11.39 | 2 | 4 | 2.39 | 0 | 0 | 0 | 1.728 | 1.806 | 1.746 | 1.423 | 1.464 | 1.436 | 0.402 | 0.412 | 0.409 |
| 3D_June | June | Control | 118 | 11 | 15 | 13.22 | 2 | 4 | 2.6 | 0 | 0 | 0 | 1.737 | 1.857 | 1.811 | 1.388 | 1.409 | 1.403 | 0.449 | 0.449 | 0.449 |
| 4D_June | June | Control | 123 | 11 | 14 | 12.19 | 0 | 1 | 0.09 | 0 | 0 | 0 | 0.871 | 0.943 | 0.891 | 0.410 | 0.431 | 0.413 | 0.805 | 0.813 | 0.812 |
| 5A_June | June | Control | 66 | 13 | 16 | 14.09 | 1 | 2 | 1.1 | 0 | 0 | 0 | 2.107 | 2.219 | 2.150 | 1.928 | 1.967 | 1.945 | 0.258 | 0.258 | 0.258 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1F_upper | June | Control | 51 | 16 | 17 | 16.08 | 2 | 4 | 2.73 | 0 | 0 | 0 | 2.379 | 2.445 | 2.383 | 2.268 | 2.347 | 2.270 | 0.196 | 0.216 | 0.215 |
| 1F_lower | June | Control | 50 | 11 | 11 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 1.742 | 1.742 | 1.742 | 1.458 | 1.458 | 1.458 | 0.360 | 0.360 | 0.360 |
| 2B_upper | June | Control | 47 | 9 | 9 | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 1.598 | 1.598 | 1.598 | 1.365 | 1.365 | 1.365 | 0.426 | 0.426 | 0.426 |
| 2B_lower | June | Control | 50 | 8 | 10 | 8.39 | 1 | 3 | 1.39 | 0 | 0 | 0 | 1.645 | 1.779 | 1.675 | 1.472 | 1.562 | 1.498 | 0.380 | 0.400 | 0.393 |
| 3D_upper | June | Control | 73 | 10 | 13 | 11.7 | 2 | 3 | 2.08 | 0 | 0 | 0 | 1.707 | 1.827 | 1.779 | 1.436 | 1.462 | 1.452 | 0.411 | 0.411 | 0.411 |
| 3D_lower | June | Control | 45 | 6 | 9 | 7.52 | 0 | 2 | 0.52 | 0 | 0 | 0 | 1.434 | 1.557 | 1.500 | 1.194 | 1.214 | 1.206 | 0.511 | 0.511 | 0.511 |
| 4D_upper | June | Control | 52 | 10 | 12 | 11.07 | 0 | 1 | 0.09 | 0 | 0 | 0 | 1.140 | 1.255 | 1.174 | 0.625 | 0.680 | 0.631 | 0.712 | 0.731 | 0.729 |
| 4D_lower | June | Control | 71 | 7 | 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.593 | 0.593 | 0.593 | 0.271 | 0.271 | 0.271 | 0.873 | 0.873 | 0.873 |
| 5A_upper | June | Control | 44 | 13 | 16 | 14.07 | 1 | 2 | 1.08 | 0 | 0 | 0 | 2.216 | 2.385 | 2.280 | 2.125 | 2.237 | 2.171 | 0.250 | 0.250 | 0.250 |
| 5A_lower | June | Control | 22 | 3 | 4 | 3.02 | 0 | 1 | 0.02 | 0 | 0 | 0 | 0.937 | 1.024 | 0.939 | 0.878 | 0.899 | 0.878 | 0.591 | 0.591 | 0.591 |
| | | | | | | | | | | | | | | | | | | | | | |
| 1E | Oct | Biocide | 54 | 9 | 9 | 9 | 2 | 2 | 2 | 0 | 0 | 0 | 0.791 | 0.791 | 0.791 | 0.367 | 0.367 | 0.367 | 0.833 | 0.833 | 0.833 |
| 2D | Oct | Biocide | 36 | 8 | 10 | 8.74 | 0 | 0 | 0 | 0 | 0 | 0 | 1.296 | 1.388 | 1.335 | 0.874 | 0.885 | 0.879 | 0.639 | 0.639 | 0.639 |
| 3C | Oct | Biocide | 52 | 11 | 14 | 12.68 | 0 | 0 | 0 | 0 | 0 | 0 | 1.673 | 1.995 | 1.890 | 1.364 | 1.573 | 1.526 | 0.423 | 0.423 | 0.423 |
| 4B | Oct | Biocide | 57 | 14 | 20 | 16.11 | 6 | 11 | 7.02 | 0 | 1 | 0.15 | 1.739 | 2.153 | 1.879 | 1.171 | 1.826 | 1.267 | 0.316 | 0.544 | 0.519 |
| 5E | Oct | Biocide | 63 | 12 | 14 | 12.75 | 0 | 0 | 0 | 0 | 0 | 0 | 1.785 | 1.966 | 1.872 | 1.527 | 1.621 | 1.579 | 0.365 | 0.365 | 0.365 |
| 1F | Oct | Control | 79 | 13 | 16 | 14.17 | 2 | 3 | 2.04 | 1 | 1 | 1 | 1.813 | 1.949 | 1.864 | 1.479 | 1.547 | 1.496 | 0.367 | 0.380 | 0.379 |
| 2B | Oct | Control | 57 | 11 | 12 | 11.09 | 0 | 1 | 0.02 | 0 | 0 | 0 | 1.623 | 1.670 | 1.625 | 1.314 | 1.326 | 1.314 | 0.421 | 0.421 | 0.421 |
| 3D | Oct | Control | 66 | 9 | 12 | 10.69 | 1 | 1 | 1 | 0 | 0 | 0 | 1.426 | 1.637 | 1.576 | 1.183 | 1.249 | 1.236 | 0.470 | 0.470 | 0.470 |
| 4D | Oct | Control | 45 | 7 | 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.946 | 0.946 | 0.946 | 0.552 | 0.552 | 0.552 | 0.756 | 0.756 | 0.756 |
| 1B | Oct | Lime | 69 | 8 | 9 | 8.71 | 0 | 1 | 0.01 | 0 | 0 | 0 | 1.121 | 1.170 | 1.155 | 0.723 | 0.729 | 0.727 | 0.681 | 0.681 | 0.681 |
| 2C | Oct | Lime | 46 | 7 | 8 | 7.43 | 0 | 1 | 0.32 | 0 | 0 | 0 | 1.354 | 1.448 | 1.394 | 1.040 | 1.110 | 1.070 | 0.543 | 0.565 | 0.556 |
| 3B | Oct | Lime | 66 | 9 | 12 | 10.66 | 1 | 1 | 1 | 0 | 0 | 0 | 1.438 | 1.637 | 1.582 | 1.117 | 1.195 | 1.179 | 0.515 | 0.515 | 0.515 |
| 4F | Oct | Lime | 58 | 13 | 14 | 13.12 | 1 | 4 | 2.38 | 0 | 0 | 0 | 1.453 | 1.492 | 1.458 | 0.838 | 0.842 | 0.839 | 0.655 | 0.655 | 0.655 |
| 5B | Oct | Lime | 79 | 12 | 14 | 13.58 | 1 | 1 | 1 | 1 | 1 | 1 | 1.774 | 2.026 | 1.964 | 1.436 | 1.677 | 1.614 | 0.329 | 0.354 | 0.330 |
| 1C | Oct | Nitrogen | 64 | 12 | 14 | 13.05 | 0 | 0 | 0 | 0 | 0 | 0 | 1.908 | 2.000 | 1.958 | 1.653 | 1.687 | 1.672 | 0.297 | 0.297 | 0.297 |
| 2A | Oct | Nitrogen | 73 | 12 | 14 | 12.97 | 1 | 2 | 1.02 | 0 | 0 | 0 | 1.811 | 1.936 | 1.904 | 1.541 | 1.642 | 1.631 | 0.356 | 0.384 | 0.358 |
| 3F | Oct | Nitrogen | 63 | 12 | 14 | 13.02 | 1 | 2 | 1.02 | 0 | 0 | 0 | 1.639 | 1.723 | 1.680 | 1.083 | 1.096 | 1.090 | 0.571 | 0.571 | 0.571 |
| 4E | Oct | Nitrogen | 75 | 14 | 15 | 14.17 | 1 | 3 | 1.2 | 0 | 0 | 0 | 1.583 | 1.609 | 1.588 | 1.039 | 1.041 | 1.039 | 0.573 | 0.573 | 0.573 |
| 5C | Oct | Nitrogen | 24 | 8 | 8 | 8 | 0 | 1 | 0.03 | 0 | 0 | 0 | 1.965 | 1.965 | 1.965 | 2.186 | 2.186 | 2.186 | 0.167 | 0.167 | 0.167 |
| 1A | Oct | Nitrogen + lime | 68 | 23 | 25 | 23.29 | 5 | 7 | 5.39 | 0 | 0 | 0 | 2.596 | 2.697 | 2.613 | 2.155 | 2.259 | 2.176 | 0.294 | 0.309 | 0.305 |
| 2E | Oct | Nitrogen + lime | 81 | 10 | 13 | 11.79 | 1 | 3 | 1.3 | 0 | 0 | 0 | 1.461 | 1.781 | 1.711 | 1.187 | 1.450 | 1.402 | 0.395 | 0.407 | 0.405 |
| 3A | Oct | Nitrogen + lime | 66 | 17 | 20 | 18.8 | 0 | 2 | 0.65 | 0 | 1 | 0.06 | 2.389 | 2.570 | 2.484 | 2.228 | 2.434 | 2.288 | 0.212 | 0.212 | 0.212 |
| 4C | Oct | Nitrogen + lime | 66 | 12 | 15 | 13.68 | 1 | 2 | 1.05 | 0 | 0 | 0 | 1.722 | 1.937 | 1.887 | 1.435 | 1.562 | 1.524 | 0.379 | 0.394 | 0.394 |
| 5D | Oct | Nitrogen + lime | 52 | 10 | 11 | 10.15 | 0 | 1 | 0.18 | 0 | 0 | 0 | 1.620 | 1.663 | 1.626 | 1.290 | 1.298 | 1.291 | 0.462 | 0.462 | 0.462 |

Table 6.3.2 Diversity data for 4bp_MOTU from all June and October plots, showing mean, maximum and minimum values from 100 runs of **define_MOTU.pl**



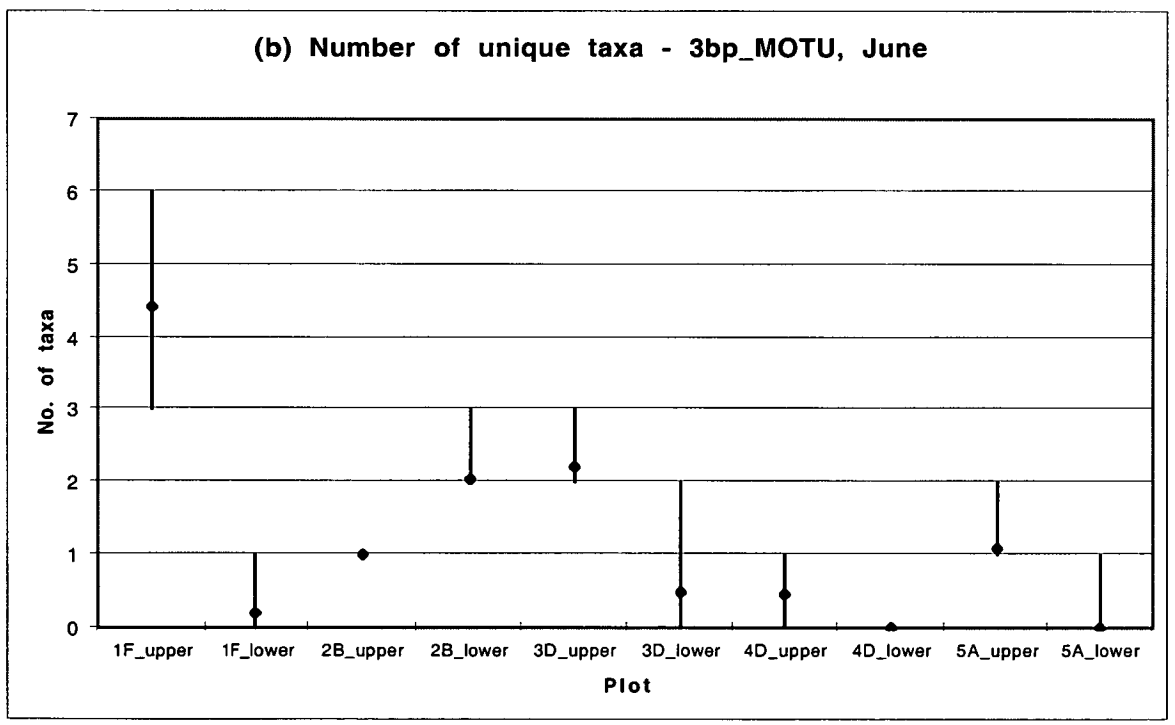
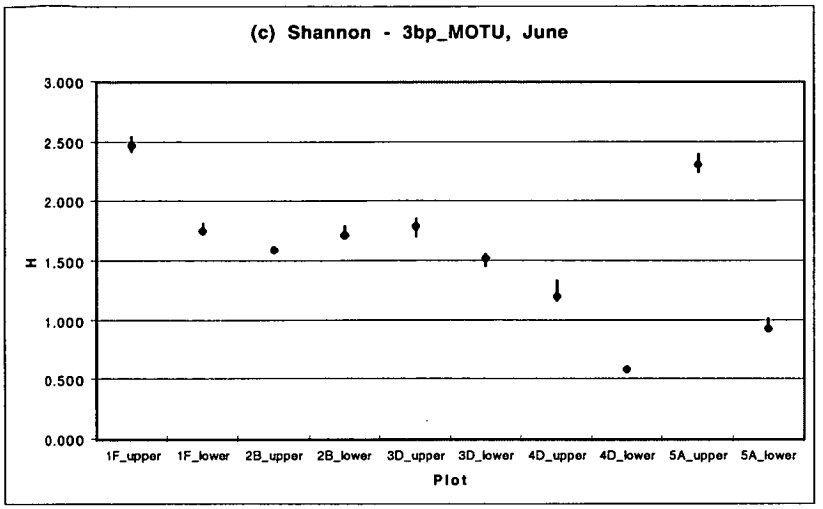
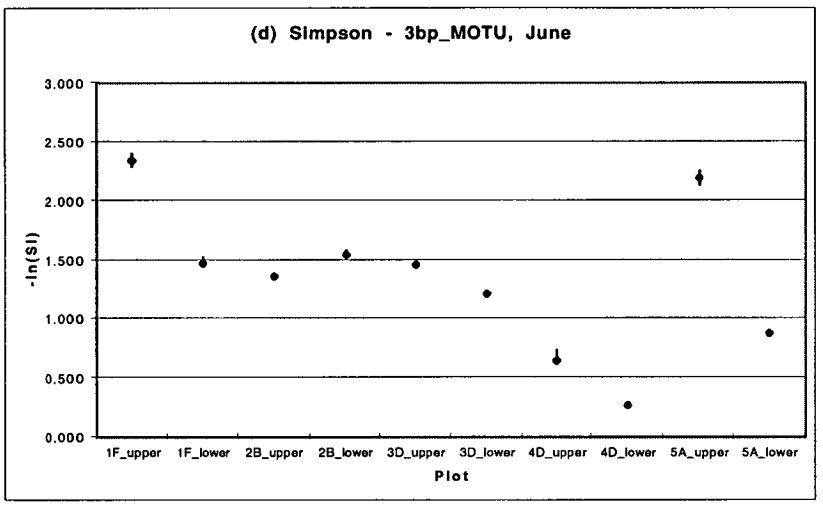




Figure 6.3.1 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the June 2001 samples for MOTU with 3bp variation, showing mean and range over 100 runs.



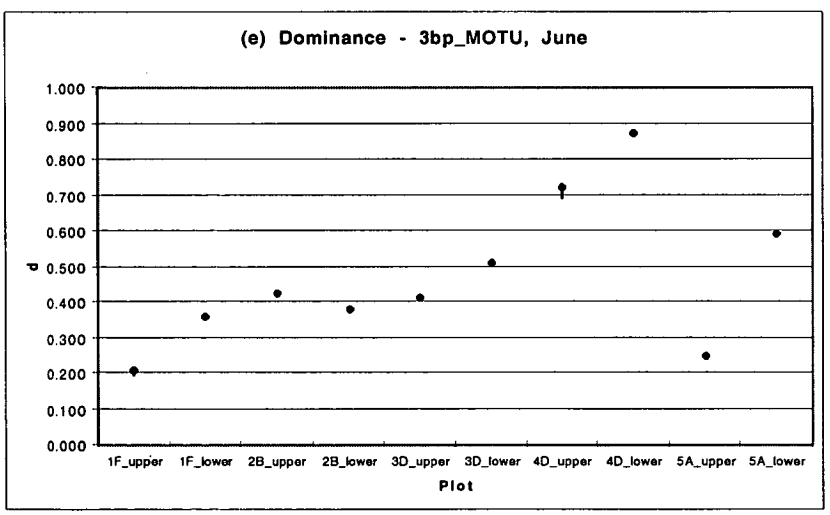
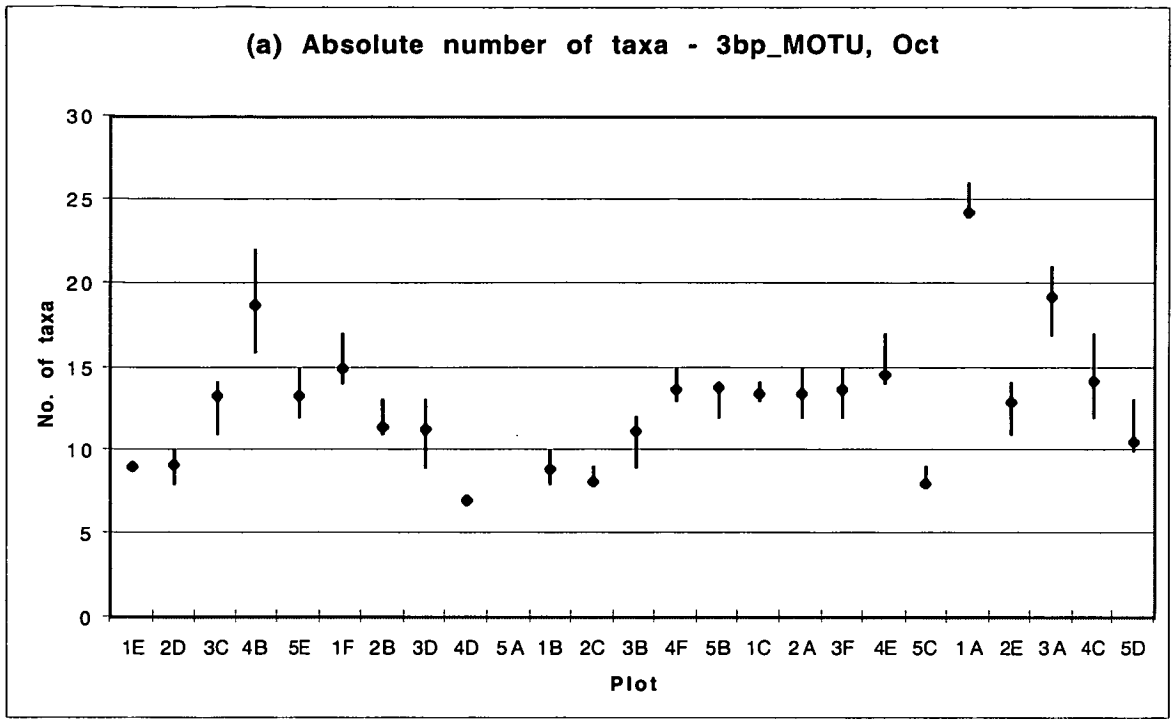




Figure 6.3.1 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the June 2001 samples for MOTU with 3bp variation, showing mean and range over 100 runs.



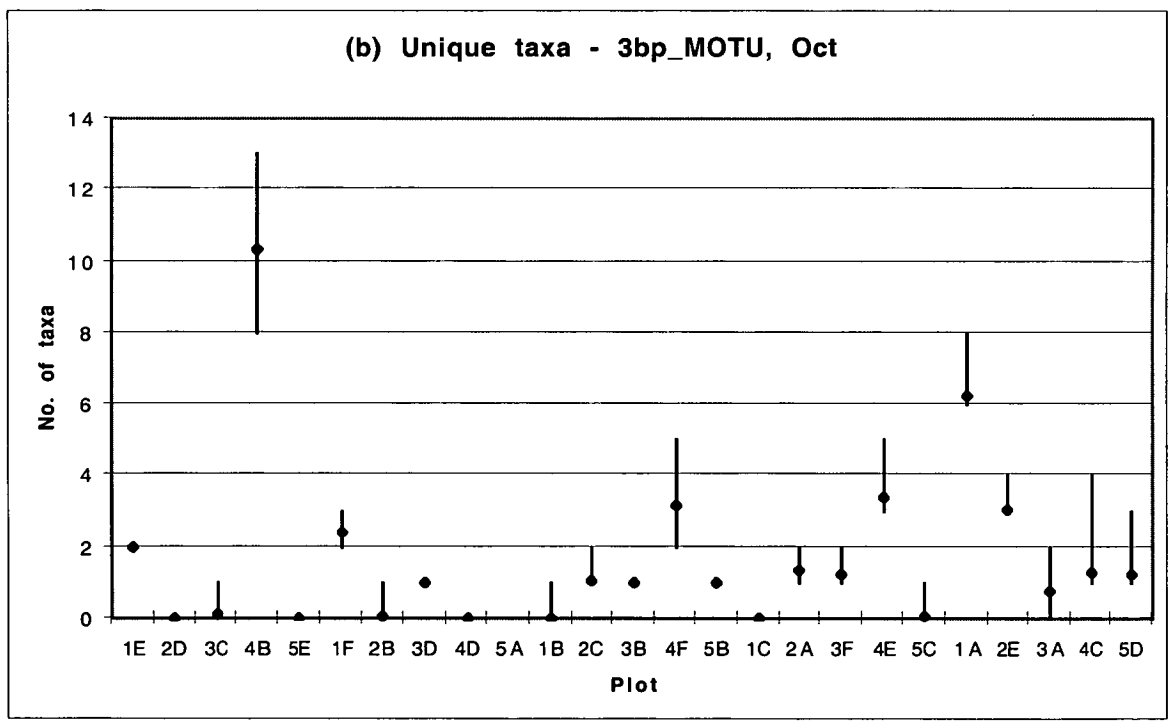
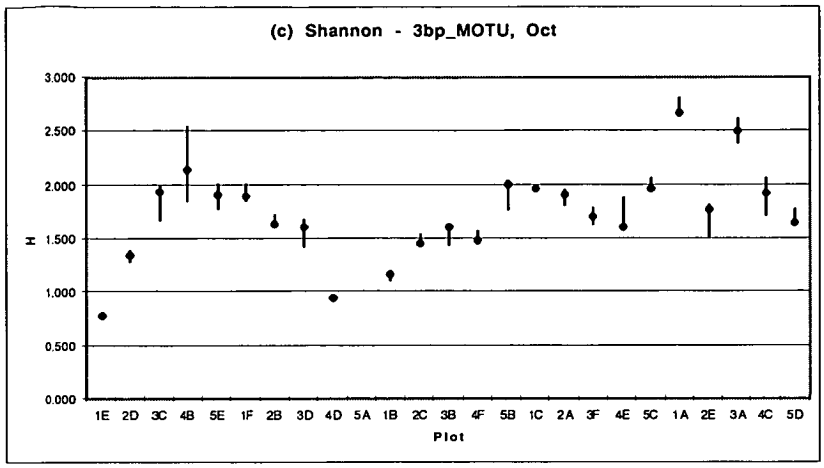
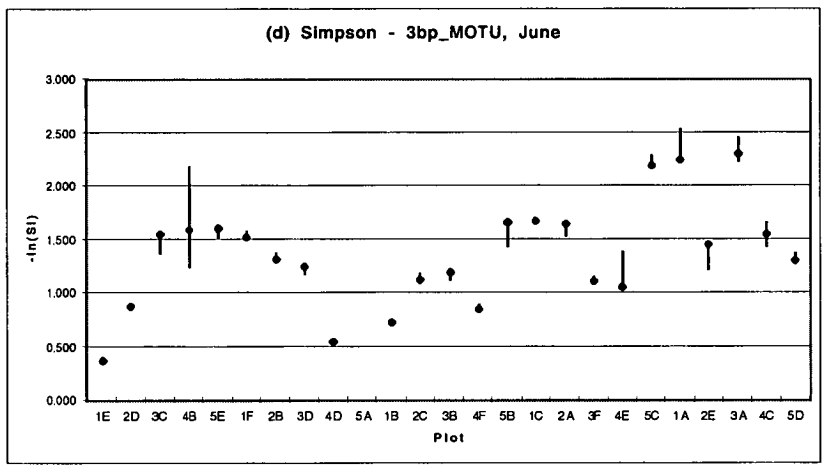




Figure 6.3.2 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the October 2001 samples for MOTU with 3bp variation, showing mean and range over 100 runs.



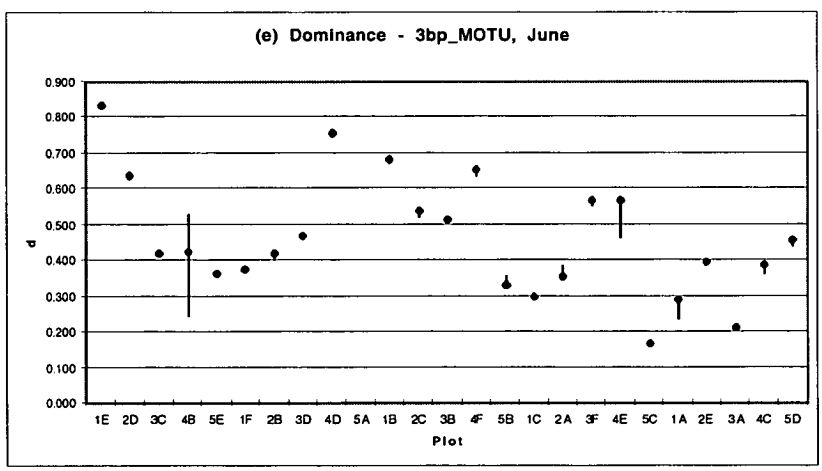
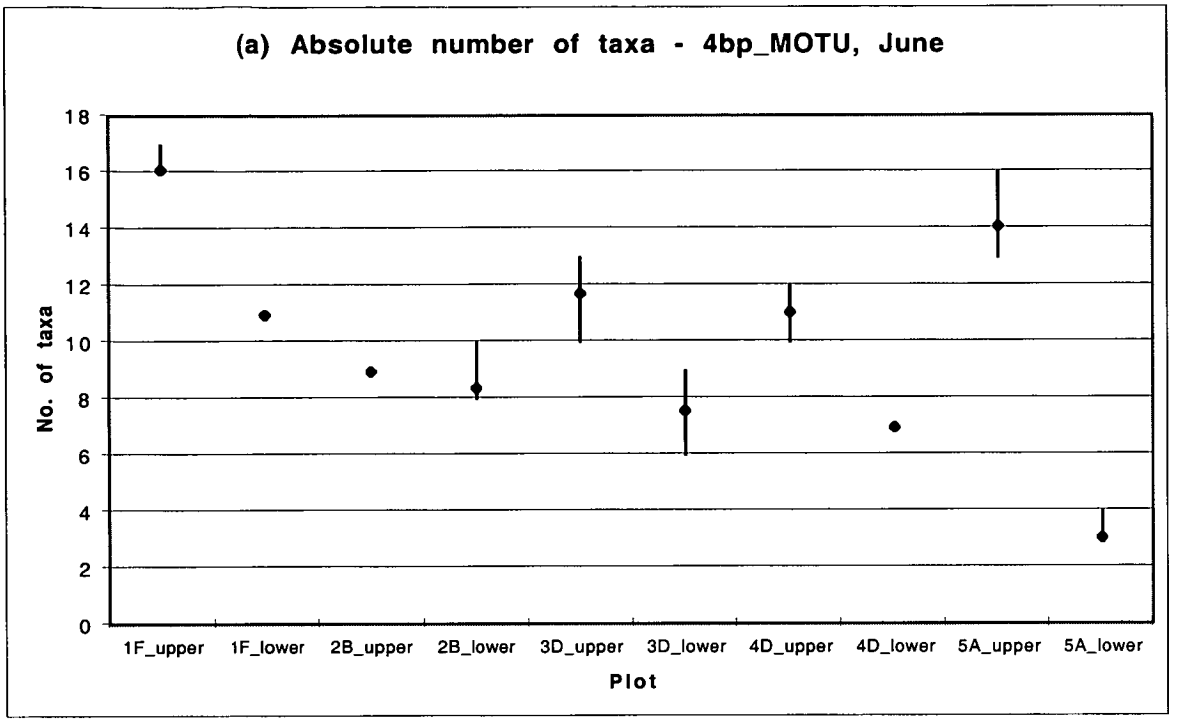




Figure 6.3.2 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the October 2001 samples for MOTU with 3bp variation, showing mean and range over 100 runs.



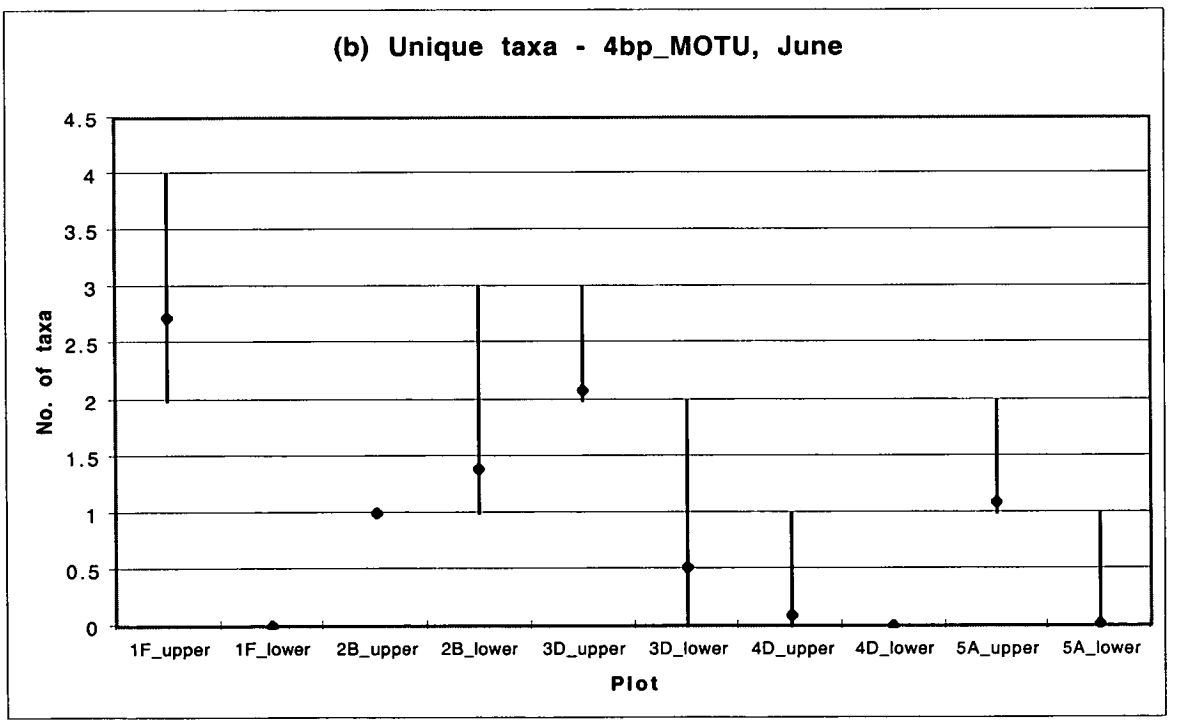
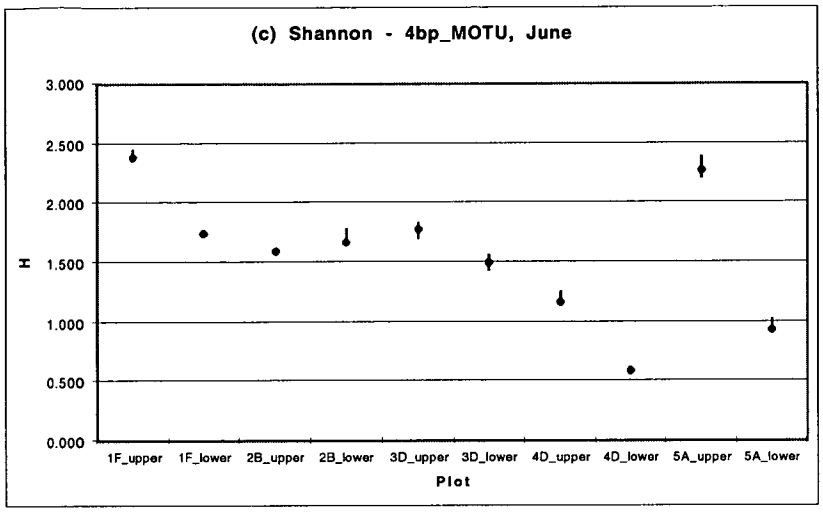
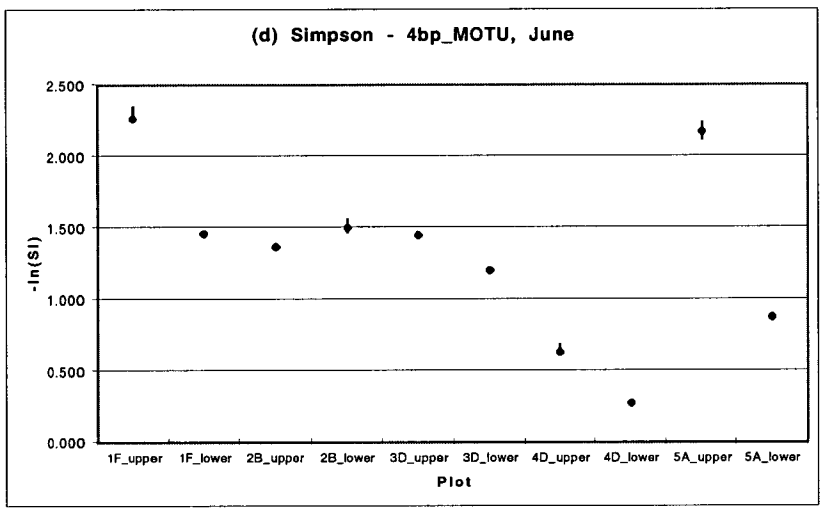




Figure 6..3.3 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the June 2001 samples for MOTU with 4bp variation, showing mean and range over 100 runs.



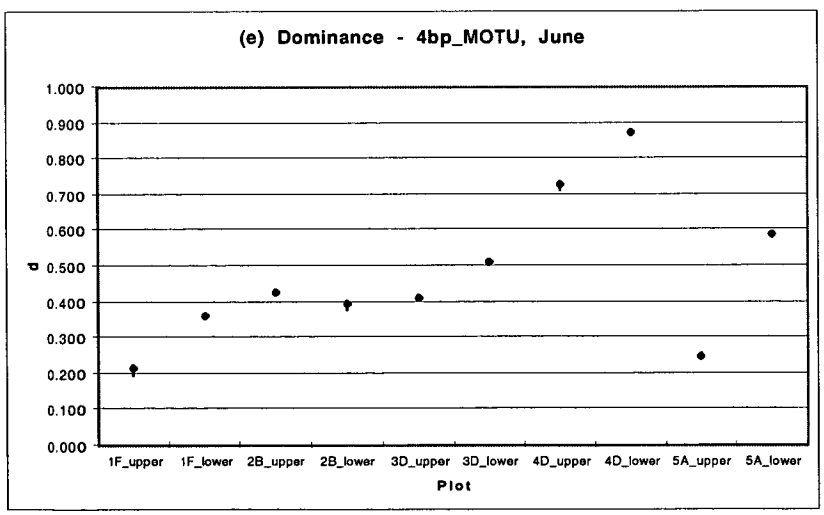
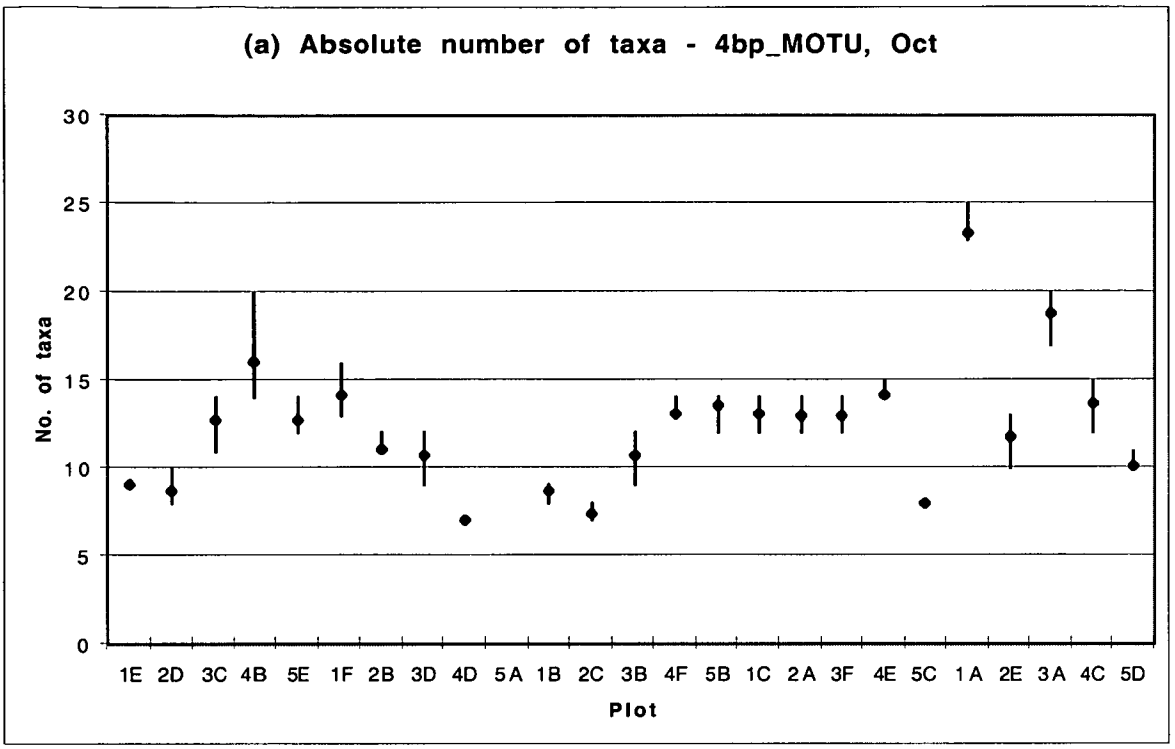




Figure 6.3.3 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the June 2001 samples for MOTU with 4bp variation, showing mean and range over 100 runs.



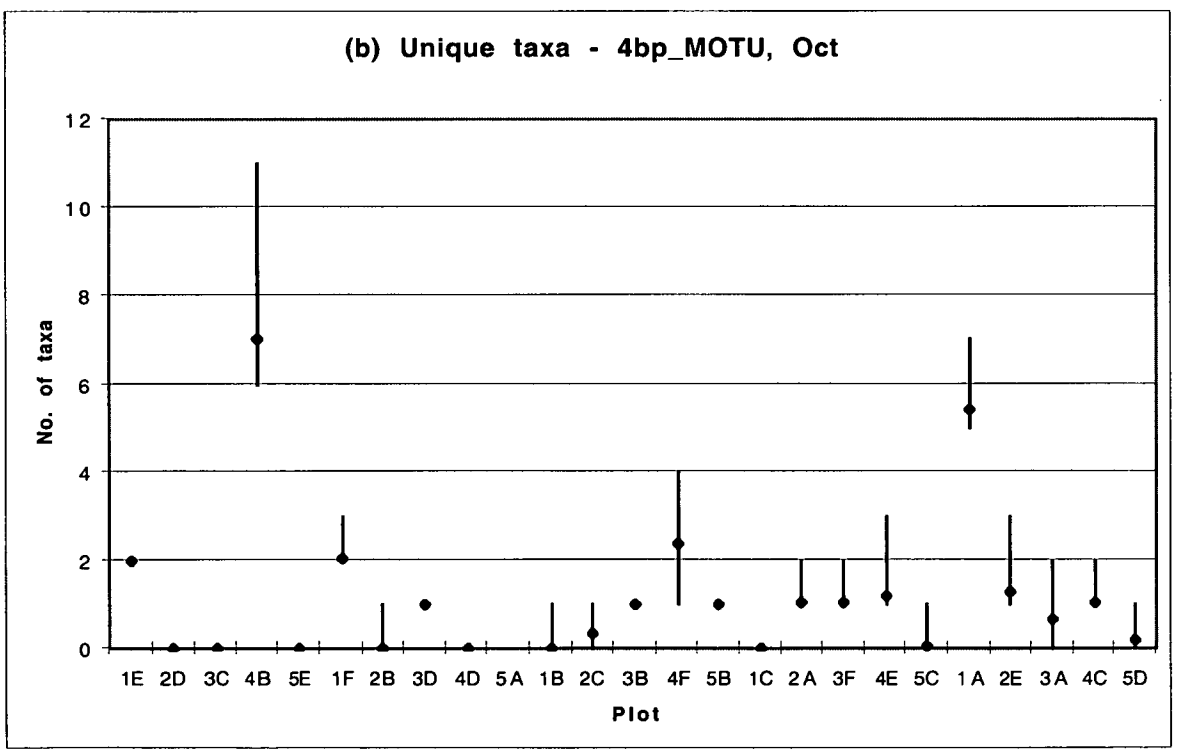
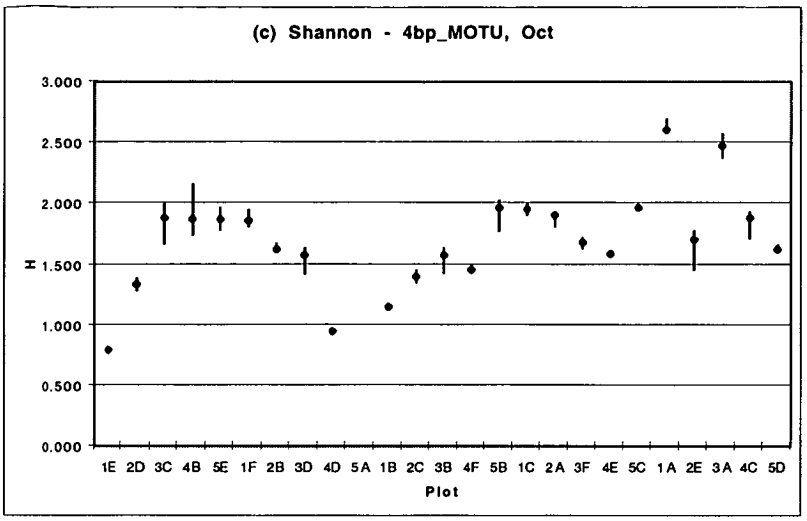
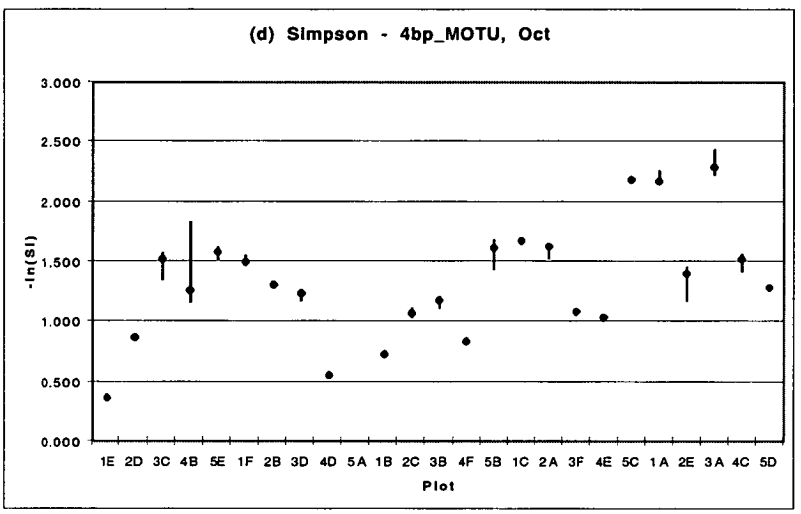




Figure 6.3.4 Graphs of (a) absolute numbers of taxa per plot, and (b) numbers of taxa unique to each plot, for the October 2001 samples for MOTU with 4bp variation, showing mean and range over 100 runs.



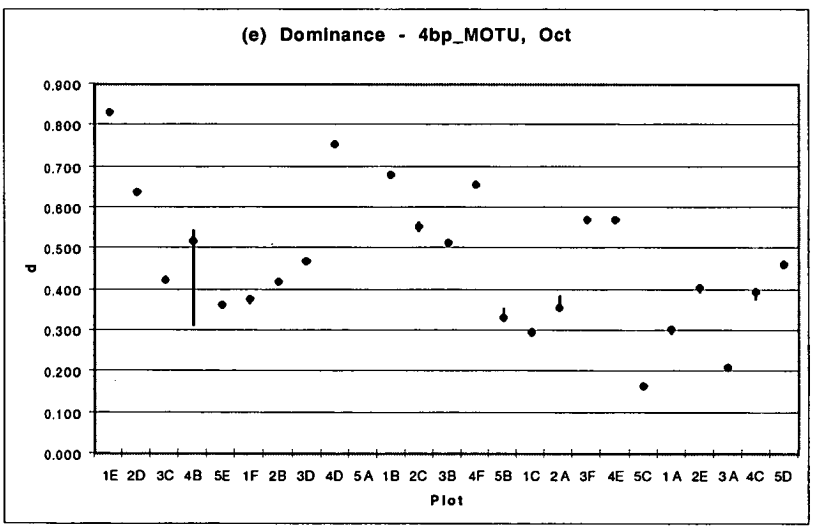




Figure 6.3.4 (cont.) Graphs of (c) Shannon, (d) Simpson, and (e) dominance index values for the October 2001 samples for MOTU with 4bp variation, showing mean and range over 100 runs.

As expected, the absolute number of taxa found, and consequently the absolute value of all the diversity measures, fell slightly as the level of variation allowed within a taxon increased. However, the relative diversities within each set of samples remains broadly the same: for example, 1A always has the greatest number of taxa and 4D the least, while 4B always contains the greatest number of unique taxa. The overall rank orders for the diversity indices are:

June samples, Shannon index:

2bp_MOTU - 1Fu,5Au,3Du,1Fl,2Bl,2Bu,3Dl,4Du,5Al,4Dl
3bp_MOTU - 1Fu,5Au,3Du,1Fl,2Bl,2Bu,3Dl,4Du,5Al,4Dl
4bp_MOTU - 1Fu,5Au,3Du,1Fl,2Bl,2Bu,3Dl,4Du,5Al,4Dl

June samples, Simpson index:

2bp_MOTU - 1Fu,5Au,2Bl,1Fl,3Du,2Bu,3Dl,5Al,4Du,4Dl
3bp_MOTU - 1Fu,5Au,2Bl,1Fl,3Du,2Bu,3Dl,5Al,4Du,4Dl
4bp_MOTU - 1Fu,5Au,2Bl,1Fl,3Du,2Bu,3Dl,5Al,4Du,4Dl

June samples, Dominance index:

2bp_MOTU - 1Fu,5Au,1Fl,2Bl,3Du,2Bu,3Dl,5Al,4Du,4Dl
3bp_MOTU - 1Fu,5Au,1Fl,2Bl,3Du,2Bu,3Dl,5Al,4Du,4Dl
4bp_MOTU - 1Fu,5Au,1Fl,2Bl,3Du,2Bu,3Dl,5Al,4Du,4Dl

Thus, for the June samples, the rank orders for the Shannon, Simpson and dominance indices are 100% identical whether MOTU are defined at the 2, 3 or 4 bp levels.

Oct samples, Shannon index:

2bp_MOTU -
1A,3A,4B,3C,5B,1F,1C,5E,5C,4C,2A,3F,2E,2B,5D,4E,3D,3B,4F,2C,2D,1B,4D,1E
3bp_MOTU -
1A,3A,4B,5B,1C,5C,3C,4C,2A,5E,1F,2E,3F,5D,2B,3B,3D,4E,4F,2C,2D,1B,4D,1E
4bp_MOTU -
1A,3A,5C,5B,1C,2A,3C,4C,4B,5E,1F,2E,3F,5D,2B,4E,3B,3D,4F,2C,2D,1B,4D,1E

Oct samples, Simpson index:

2bp_MOTU -

3A,1A,5C,4B,1C,5B,2A,3C,5E,1F,4C,2E,2B,5D,3D,3B,3F,2C,4E,4F,2D,1B,4D,1E

3bp_MOTU -

3A,1A,5C,1C,5B,2A,5E,4B,3C,4C,1F,2E,2B,5D,3D,3B,2C,3F,4E,2D,4F,1B,4D,1E

4bp_MOTU -

3A,5C,1A,1C,2A,5B,5E,3C,4C,1F,2E,2B,5D,4B,3D,3B,3F,2C,4E,2D,4F,1B,4D,1E

Oct samples, Dominance index:

2bp_MOTU -

5C,3A,1A,1C,4B,5B,2A,5E,1F,2B,4C,2E,3C,5D,3D,3B,2C,3F,4E,4F,2D,1B,4D,1E

3bp_MOTU -

5C,3A,1A,1C,5B,2A,5E,1F,4C,2E,2B,3C,4B,5D,3D,3B,2C,3F,4E,2D,4F,1B,4D,1E

4bp_MOTU -

5C,3A,1C,1A,5B,2A,5E,1F,4C,2E,2B,3C,5D,3D,3B,4B,2C,3F,4E,2D,4F,1B,4D,1E

For October, with its larger number of samples, some inconsistencies in the values of the indices between MOTU sets are seen, though the plots at the beginning and end of the series remain approximately the same, with the variations seen in the intermediate plots. Plot 4B, in particular, is very inconsistent in its rank order position, but this is not unexpected given the "flock" of closely-related taxa it has previously been found to contain.

It is also notable that the variability in taxon number across 100 runs (i.e. the size of the error bars on the graphs) changes rather little when the MOTU designation threshold is increased to 3 and 4 bases. In most cases the range remains of approximately the same magnitude, and the plots which previously show the widest range of values (such as 4B) still show the widest range at the higher levels. This shows that the influence of processing order on taxon assignment does not disappear when more variation is allowed within a taxon, but is likely to remain an issue whatever level is chosen.

## 6.4 Taxonomic diversity

Since each MOTU can also be assigned to a known taxonomic group, it is also possible to examine the distributions of particular groups of nematodes. Only one set of MOTU was used in this analysis, as variations in MOTU number should not affect the higher taxonomic groups to which sequences belong. The run already discussed in Chapter 4, with a total of 140 MOTU, was used. On the basis of similarity to known sequences (see Figure 4.7.1), each MOTU was assigned a taxonomic name in the SQL database: these are listed in Appendix 2. Some could be identified to genus level, some to family, and

others only to order/suborder. The classification scheme of De Ley and Blaxter (2001) was followed, in which the nematodes previously placed in the order Tylenchida (Thorne 1949; Siddiqi 2000) are reduced to the rank of a suborder (Tylenchina) within the order Rhabditida. Tylenchina as used here contains Panagrolaimomorpha (including Steinernematidae), Cephalobomorpha and Tylenchomorpha, while the suborder Rhabditina contains Bunonematomorpha, Diplogasteromorpha and Rhabditomorpha.

Those MOTU which were known to family level could also be assigned coloniser-persistor values, allowing the maturity index (MI) of each plot to be calculated. Table 6.4.1 shows the number of members of each of the eight major nematode groups recorded in this survey (Tylenchina, Dorylaimida, Enoplida, Mononchida, Rhabditina, Plectida, Chromadorida and Monhysterida), along with the MI values from each plot.

| Sample | Date | Treatment | No. Indiv. | MI | Tylenchina | Dorylaimida | Enoplida | Mononchida | Rhabditina | Plectida | Chromadorida | Monhysterida |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| 1F | June | Control | 101 | 3.781 | 27 | 36 | 18 | 18 | 1 | 1 | | |
| 2B | June | Control | 97 | 3.639 | 40 | 30 | 23 | 3 | 1 | | | |
| 3D | June | Control | 118 | 3.396 | 55 | 36 | 24 | 3 | | | | |
| 4D | June | Control | 123 | 3.140 | 100 | 13 | 3 | 2 | 4 | 1 | | |
| 5A | June | Control | 66 | 3.661 | 18 | 27 | 16 | 3 | | 2 | | |
| | | | | | | | | | | | | |
| 1F_upper | June | Control | 51 | 3.612 | 10 | 15 | 15 | 10 | | 1 | | |
| 1F_lower | June | Control | 50 | 3.957 | 17 | 21 | 3 | 8 | 1 | | | |
| 2B_upper | June | Control | 47 | 3.644 | 20 | 13 | 11 | 2 | 1 | | | |
| 2B_lower | June | Control | 50 | 3.632 | 20 | 17 | 12 | 1 | | | | |
| 3D_upper | June | Control | 73 | 3.344 | 30 | 20 | 20 | 3 | | | | |
| 3D_lower | June | Control | 45 | 3.486 | 25 | 16 | 4 | | | | | |
| 4D_upper | June | Control | 52 | 3.191 | 38 | 6 | 2 | 2 | 3 | 1 | | |
| 4D_lower | June | Control | 71 | 3.104 | 62 | 7 | 1 | | 1 | | | |
| 5A_upper | June | Control | 44 | 3.675 | 5 | 18 | 16 | 3 | | 2 | | |
| 5A_lower | June | Control | 22 | 3.632 | 13 | 9 | | | | | | |
| | | | | | | | | | | | | |
| 1E | Oct | Biocide | 54 | 3.058 | 47 | 4 | 1 | 2 | | | | |
| 2D | Oct | Biocide | 36 | 3.333 | 25 | 10 | | 1 | | | | |
| 3C | Oct | Biocide | 52 | 3.460 | 23 | 21 | 2 | 2 | 1 | 1 | | 2 |
| 4B | Oct | Biocide | 57 | 3.315 | 35 | 12 | 3 | 5 | 1 | | | 1 |
| 5E | Oct | Biocide | 63 | 4.070 | 15 | 39 | 5 | 2 | 1 | | | 1 |
| 1F | Oct | Control | 79 | 3.750 | 31 | 36 | 5 | 6 | | 1 | | |
| 2B | Oct | Control | 57 | 3.706 | 26 | 25 | 4 | 1 | 1 | | | |
| 3D | Oct | Control | 66 | 3.667 | 32 | 31 | 1 | 1 | | | | 1 |
| 4D | Oct | Control | 45 | 3.250 | 34 | 6 | 3 | 2 | | | | |
| 1B | Oct | Lime | 69 | 3.377 | 48 | 15 | 4 | | 1 | | 1 | |
| 2C | Oct | Lime | 46 | 3.512 | 26 | 17 | 2 | 1 | | | | |
| 3B | Oct | Lime | 66 | 3.476 | 34 | 25 | 3 | 3 | | 1 | | |
| 4F | Oct | Lime | 58 | 2.964 | 43 | 3 | 2 | 5 | 3 | 1 | 1 | |
| 5B | Oct | Lime | 79 | 3.570 | 27 | 36 | 2 | 8 | 2 | 4 | | |
| 1C | Oct | Nitrogen | 64 | 3.763 | 20 | 28 | 5 | 5 | 4 | 1 | | 1 |
| 2A | Oct | Nitrogen | 73 | 3.957 | 10 | 45 | 5 | 12 | 1 | | | |
| 3F | Oct | Nitrogen | 63 | 3.397 | 37 | 13 | 4 | 9 | | | | |
| 4E | Oct | Nitrogen | 75 | 3.338 | 46 | 18 | 3 | | 6 | 2 | | |
| 5C | Oct | Nitrogen | 24 | 3.591 | 9 | 8 | 6 | 1 | | | | |
| 1A | Oct | Nitrogen + lime | 68 | 3.220 | 30 | 15 | 8 | 8 | 5 | 2 | | |
| 2E | Oct | Nitrogen + lime | 81 | 3.538 | 34 | 37 | 5 | 3 | | 2 | | |
| 3A | Oct | Nitrogen + lime | 66 | 3.296 | 25 | 22 | 5 | 8 | 3 | 3 | | |
| 4C | Oct | Nitrogen + lime | 66 | 3.734 | 27 | 29 | 4 | 5 | | 1 | | |
| 5D | Oct | Nitrogen + lime | 52 | 3.633 | 26 | 20 | 3 | 2 | 1 | | | |

Table 6.4.1 Distribution of nematodes by major taxonomic group

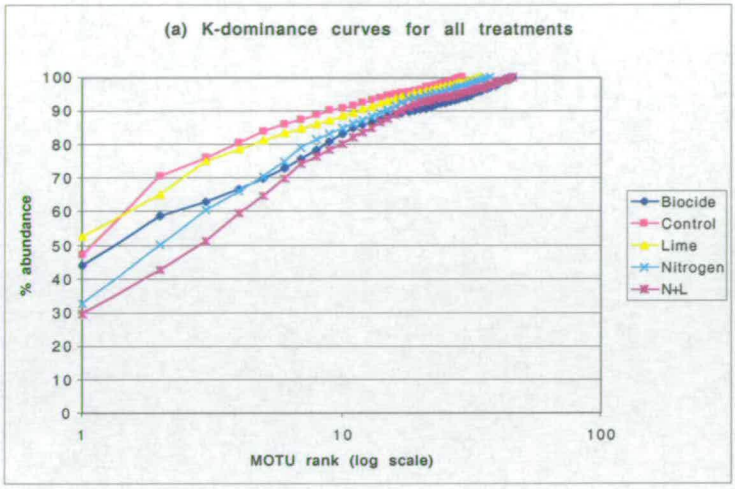
The maturity index changes very little across the site, always remaining around 3-4, suggesting that the site as a whole constitutes a relatively stable ecosystem which has not been recently disturbed (Bongers 1990). This is due to the fact that all plots are dominated by either *Helicotylenchus* (3, intermediate on the c-p scale) or *Aporcelaimellus* (5, an extreme persister); groups with low values, such as Rhabditidae and Cephalobidae, are occasionally present but are never dominant. It appears that the MI is not a sensitive indicator of changes in diversity at this site. Not even the biocide treatment, which might be expected to disturb the nematode community, has had a noticeable effect.

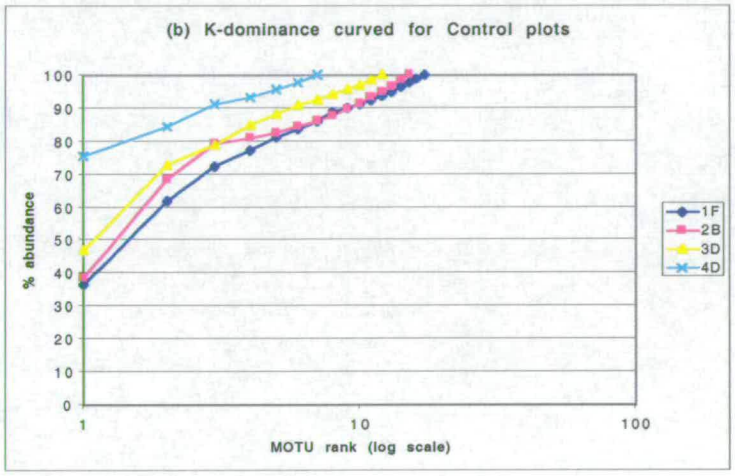
The taxonomic classification shows that tylenchs and dorylaims are the most abundant overall, but plots differ in which of the two groups is dominant.

## 6.5 K-Dominance curves

K-dominance curves are another common method for visualising patterns in diversity (e.g. Eyualem et al. 2001). These are made by plotting the cumulative percentage abundance of each taxon against its rank order on a log scale. Figure 6.5.1 (a) shows a set of combined curves for all treatments, while parts (b)-(f) show each treatment individually, with a curve for each plot within a treatment.

In general, lines low down on the graph represent higher diversity (more taxa and more even distribution of taxa) while lines higher up represent lower diversity. The intercept on the y-axis represents the percentage abundance of the dominant taxon. Overall, it can be seen from graph (a) that the control plots are less diverse than the nitrogen, line or nitrogen+lime plots, but are approximately the same as the lime plots. However, the remaining graphs show that there is considerable variability between plots within a treatment.



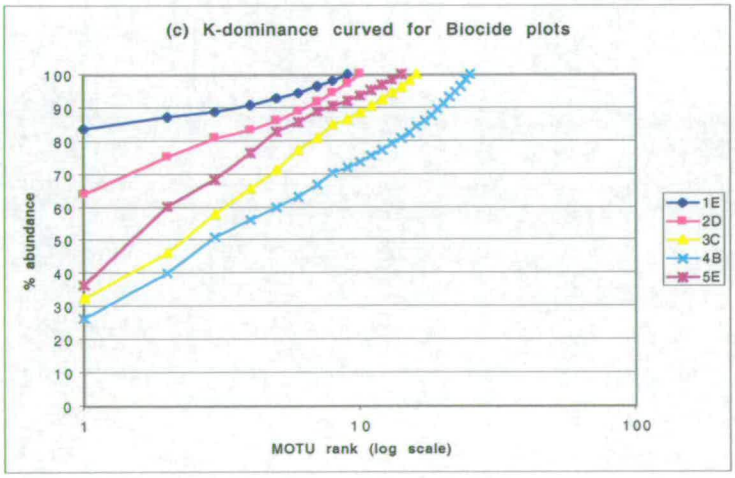
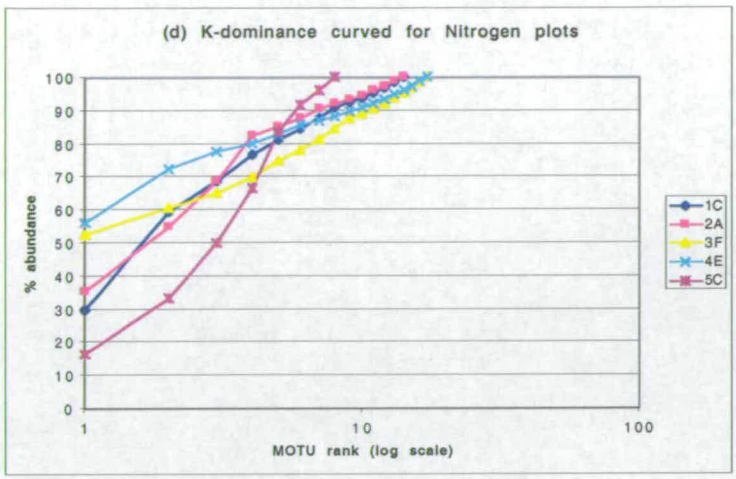
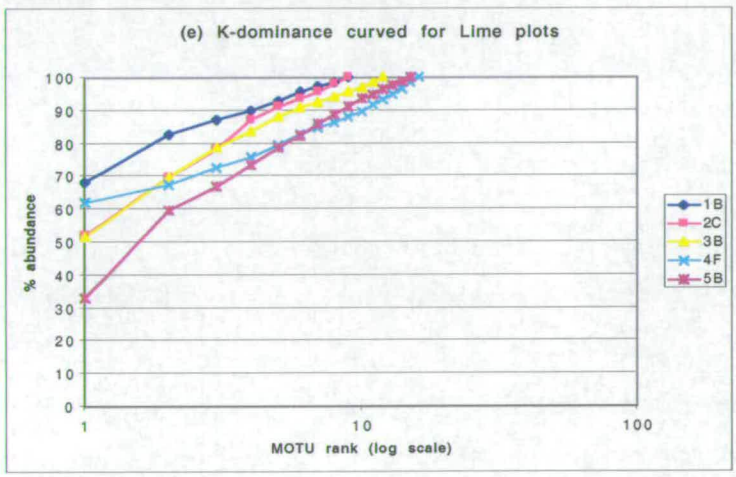




Figure 6.5.1 K-dominance curves for (a) all treatments together, (b) Control plots, and (c) Biocide plots.



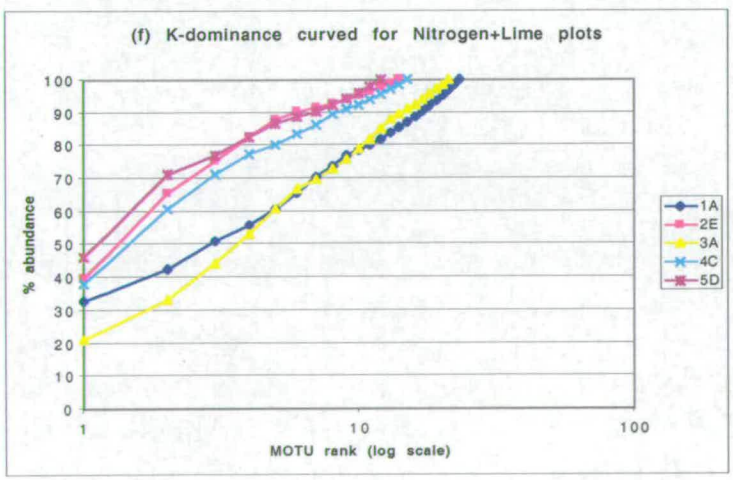




Figure 6.5.1 (cont.) K-dominance curves for (d) Nitrogen plots, (e) Lime plots, and (f) Nitrogen+Lime plots.

## 6.6 Comparison with morphological survey

For the June samples, a parallel morphological survey was carried out on a set of nematodes to genus level, allowing both sets of data to be directly compared.

### 6.6.1 Diversity indices

Using the list of genera identified and their abundances in each plot, diversity indices were calculated using the same **div_indices.pl** script as for the MOTU dataset. Results are shown in Table 6.6.1, including data both for plots split into upper and lower horizons and pooled together. Graphs for the three diversity indices are plotted in Figure 6.6.1.

It can be seen that both the morphological and molecular surveys show some common features – both agree that plot 4D_lower has the lowest diversity and 1F_upper the highest, by all three indices (though in the morphological survey 1F_upper is only fractionally higher than 5A_upper), and both usually agree that the upper horizon is more diverse than the corresponding lower horizon (except for 2B, where the MOTU survey finds a slightly higher diversity for the lower horizon).

| Sample | No. indiv. | No. taxa | Shannon | Simpson | Dominance |
|---|---|---|---|---|---|
| | | | | | |
| 1F | 191 | 18 | 2.021 | 1.609 | 0.382 |
| 2B | 190 | 17 | 1.818 | 1.394 | 0.400 |
| 3D | 192 | 15 | 2.105 | 1.858 | 0.250 |
| 4D | 191 | 19 | 1.672 | 1.005 | 0.592 |
| 5A | 192 | 19 | 2.045 | 1.569 | 0.391 |
| | | | | | |
| 1F_upper | 96 | 15 | 2.127 | 1.857 | 0.292 |
| 1F_lower | 95 | 10 | 1.654 | 1.310 | 0.474 |
| 2B_upper | 94 | 15 | 1.958 | 1.617 | 0.351 |
| 2B_lower | 96 | 11 | 1.527 | 1.185 | 0.448 |
| 3D_upper | 96 | 10 | 2.026 | 1.968 | 0.219 |
| 3D_lower | 96 | 13 | 1.947 | 1.658 | 0.323 |
| 4D_upper | 95 | 17 | 1.958 | 1.391 | 0.474 |

Table 6.6.1 Diversity indices derived from morphological survey of June samples

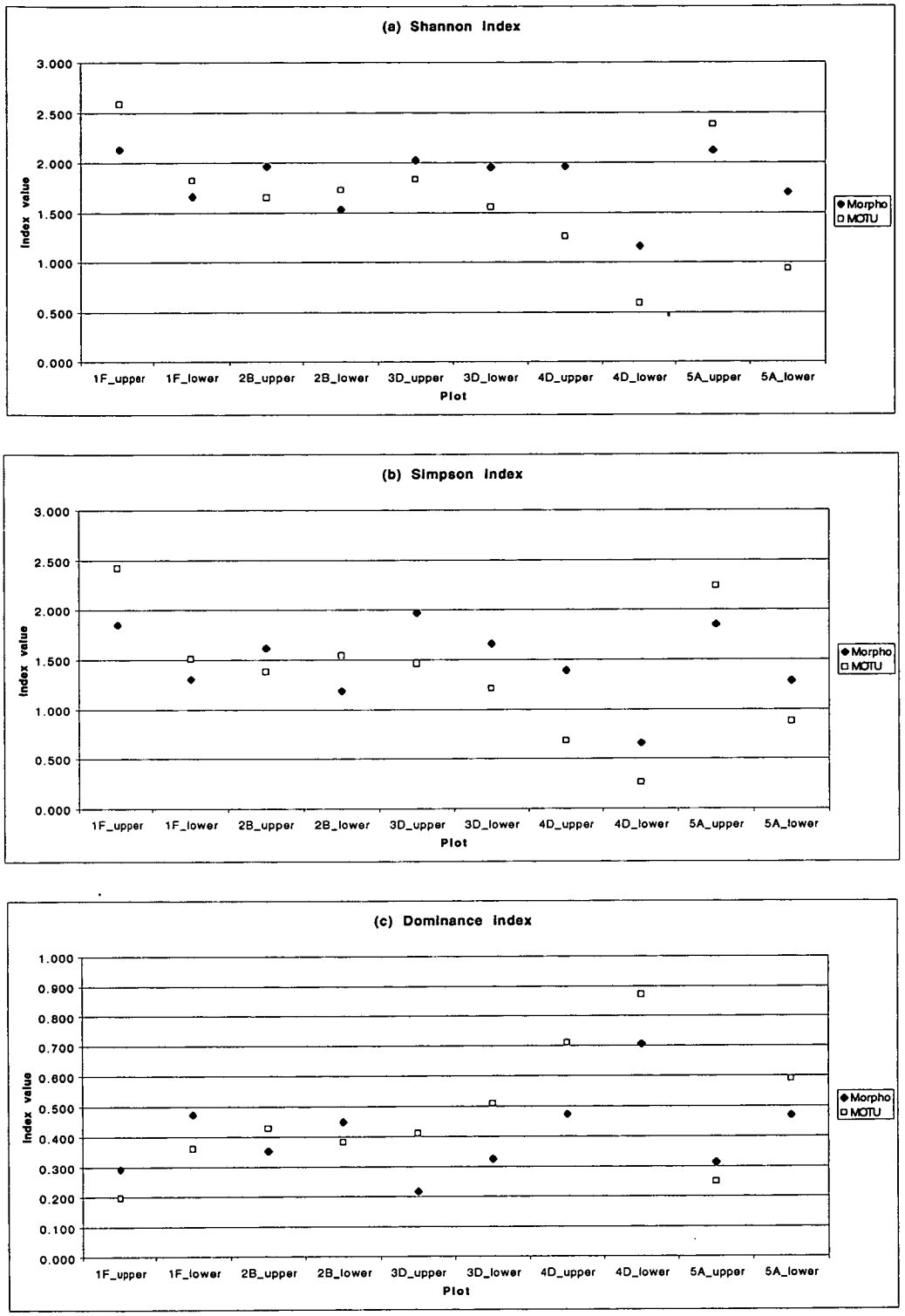



Figure 6.6.1 Graphs comparing morphological and molecular surveys for (a) Shannon, (b) Simpson and (c) dominance indices.

### 6.6.2 Taxonomic diversity

Table 6.6.2 shows a comparison of the major taxonomic groups found in the morphological and molecular surveys. Since the sample sizes differed between the two surveys, both numbers and percentages as a fraction of sample size are given. The two surveys are more different than might have been expected. In general, the molecular survey has found more tylenchs, more enoplids and fewer dorylaims than the morphological survey.

As described in Chapter 2, each soil extract was split in two, with half morphologically identified and half sequenced. Every effort was made to sample individual nematodes randomly in each case. These results, therefore, raise the possibility of a bias in PCR and/or sequencing, with tylenchs perhaps showing a greater success rate than dorylaims. However, this pattern is not always seen (1F and 5A both show more dorylaims than tylenchs in the MOTU survey, and in 4D both surveys are in approximate agreement on the relative abundances). Enoplids, too, are sometimes found to be more abundant in the MOTU survey, but not always (4D shows the reverse), indicating that it is not a straightforward case of PCR/sequencing bias, or at least that any such bias does not occur in a consistent fashion.

Each of the two surveys involved a step in which a number of nematodes was sampled for identification from among a far larger number of available specimens. Where the underlying population is complex and heterogeneous, even two entirely random samplings would not be expected to produce identical results. Here, there is the additional confounding factor of a human observer selecting which individuals are sampled; even though efforts were made to do so randomly, worker bias cannot be excluded, and different workers will not sample in precisely the same way. All of the specimens in the morphological survey were picked by Dr Eyualem Abebe; the majority of those in the molecular survey were picked by me, with the remainder by Dr Eyualem and Mark Welsh. Unintended differences in sampling technique could account for a degree of variation in the nematodes sampled from the same underlying community.

Practical differences between the two approaches could also have some effect, for example the fact that, when making slides for identification it is necessary for similar-sized nematodes to be together on the same slide (A. Eyualem, pers. comm.), which could also have introduced a degree of bias. Overall, both methods carry the potential for particular biases in their representation of the underlying community, which will tend to pull the findings of the two in different directions.

| Plot | Taxon | Numbers | | Percentages | |
|---|---|---|---|---|---|
| | | Morpho. | MOTU | Morpho. | MOTU |
| | | | | | |
| 1F | Dorylaimida | 100 | 36 | 35.64 | 52.36 |
| | Tylenchina | 43 | 27 | 26.73 | 22.51 |
| | Mononchida | 28 | 18 | 17.82 | 14.66 |
| | Enoplida | 18 | 18 | 17.82 | 9.42 |
| | Plectida | 1 | 1 | 0.99 | 0.52 |
| | Rhabditina | 1 | 1 | 0.99 | 0.52 |
| | *Total* | *191* | *101* | | |
| | | | | | |
| 2B | Dorylaimida | 108 | 30 | 30.93 | 56.84 |
| | Tylenchina | 54 | 40 | 41.24 | 28.42 |
| | Mononchida | 5 | 3 | 3.09 | 2.63 |
| | Enoplida | 20 | 23 | 23.71 | 10.53 |
| | Plectida | 3 | 0 | 0 | 1.58 |
| | Rhabditina | 0 | 1 | 1.03 | 0 |
| | *Total* | *190* | *97* | | |
| | | | | | |
| 3D | Dorylaimida | 125 | 36 | 30.51 | 65.1 |
| | Tylenchina | 55 | 55 | 46.61 | 28.65 |
| | Mononchida | 1 | 3 | 2.54 | 0.52 |
| | Enoplida | 11 | 24 | 20.34 | 5.73 |
| | *Total* | *192* | *118* | | |
| | | | | | |
| 4D | Dorylaimida | 33 | 13 | 10.57 | 17.28 |
| | Tylenchina | 140 | 100 | 81.3 | 73.3 |
| | Mononchida | 4 | 2 | 1.63 | 2.09 |
| | Enoplida | 13 | 3 | 2.44 | 6.81 |
| | Plectida | 1 | 1 | 0.81 | 0.52 |
| | Rhabditina | 0 | 4 | 3.25 | 0 |
| | *Total* | *191* | *123* | | |
| | | | | | |
| 5A | Dorylaimida | 84 | 27 | 40.91 | 43.75 |
| | Tylenchina | 88 | 18 | 27.27 | 45.83 |
| | Mononchida | 4 | 3 | 4.55 | 2.08 |
| | Enoplida | 14 | 16 | 24.24 | 7.29 |
| | Plectida | 2 | 2 | 3.03 | 1.04 |
| | *Total* | *192* | *66* | | |
| | | | | | |
| Totals | Dorylaimida | 450 | 142 | 28.12 | 47.07 |
| | Tylenchina | 380 | 240 | 47.52 | 39.75 |
| | Enoplida | 76 | 84 | 16.63 | 7.95 |
| | Mononchida | 42 | 29 | 5.74 | 4.39 |
| | Plectida | 7 | 4 | 0.79 | 0.73 |
| | Rhabditina | 1 | 6 | 1.19 | 0.1 |
| | *Total* | *956* | *505* | | |

Table 6.6.2 Comparison of morphological and molecular surveys for the June samples classified to order/suborder level, giving both numbers and percentage of sample size.

# 7. Patterns in Diversity

It has been shown how, from the basic parameters of a set of taxa and their abundances in a given sample, it is possible to derive a series of different measures, each of which reflect particular aspects of the biological diversity of the sample. These measures include the absolute number of taxa, number of unique taxa, the Shannon, Simpson and dominance indices, as well as the proportionate abundances of particular taxa of interest. It was also found that all of these values show a wide range of variation between the different plots sampled. It is therefore possible to search for patterns in this variation, or relationships between any of these measures and other parameters of interest, such as the experimental treatments which were applied, or other environmental properties of the field site. It is by this approach that we may identify the ecological factors which drive changes in diversity.

## 7.1 Variation in time - comparison of June and October samples

For four of the control plots (1F, 2B, 3D and 4D), samples were taken at two different dates, in June and October (the October sample for the fifth control plot, 5A, was lost, and so no data are available for this sample). An issue is that the June samples were larger than the October samples (in terms of numbers of individuals), due to the fact that in June both upper and lower horizons were taken separately, therefore when the two horizons are combined the June samples are, on average, roughly twice as large as the corresponding October samples (mean sample size ± st. dev - June: 101 ± 22.44; Oct: 60.79 ± 13.74). Therefore, in comparing the two dates it is important only to use measures which are sample-size independent: the diversity indices should provide appropriate measures, but number of taxa, for example, would not, as it is clearly biased by sample size (other things being equal, a larger sample is expected to contain more taxa than a smaller one from the same underlying community). Also included as variables in this analysis were the abundances of the four most common taxonomic groups found (Tylenchina, Dorylaimida, Enoplida and Mononchida), each taken as a percentage of the overall sample size (as using absolute numbers of these taxa per plot would also be biased by sample size).

It was found that none of the diversity indices shows a significant difference in mean value between the June and October samples for the 2bp_MOTU (2-sample t-test, $p > 0.05$ for all indices). This suggests that sampling date does not have a significant effect on overall nematode diversity, at least for the dates and plots compared (though, given the small sample sizes - only four observations for each date - any effect would have to be large to register as significant). None of the taxonomic groups tested showed significant variation in percentage abundance between dates, although enoplids came close to showing significance (2-sample t-test, $p=0.12$), being more abundant in all of the June plots with the exception of 4D. It is interesting to note that if the data point for 4D in June was excluded from the analysis then the

difference was found to be significant (p=0.0021). Also, if the results for plot 5A in June are included, even though there is no October value for comparison, the difference in enoplid abundance between the two groups again becomes significant (p=0.043). It is possible that there is some effect of time on this group of nematodes, but that it is not sufficiently strong or general to be unambiguously detected from this small number of data points.

## 7.2 Variation due to soil treatment and spatial position

The data from all of the October samples were analysed to determine if the experimental soil treatments had any effect on nematode diversity. However, it is important to consider that natural environmental heterogeneity on the field site may also have some effect on the nematodes. For example, since the site is on a hill there may be many subtle differences in soil microclimate between the top and bottom of the slope. The experimental design has attempted to solve this by replicating the treatments in rows from top to bottom. But there may also possibly be an effect of horizontal position from left to right across the field. Since each plot occupies a unique physical location on the field it is not possible to analyse the effects of treatment separately from spatial position; both must be considered together.

Since each plot occupies both a particular row and column, both the vertical (row 1, 2, 3, 4 or 5) and horizontal (column A, B, C, D, E or F) positions of each plot can be included as factors in the analysis to determine if nematode diversity is influenced by spatial position. Also included were measures of volumetric moisture content (%), pH and above-ground biomass (AGB) taken from the relevant plots, from data published on the Soil Biodiversity Programme website (http://mwnta.nmw.ac.uk/soilbio/data.htm). It should be noted that these data were collected by other groups within the Programme and were not taken at the same times or from the exact same within-plot locations as the nematode surveys. Additionally, sample size was included as a predictor variable to see if it made any difference to any value measured.

A general linear model (GLM) approach was taken to determine which factors are important in determining nematode diversity. For each response variable considered, the analysis began with a maximal model, with all factors included:

Response variable = treatment + vertical position + horizontal position + %moisture + pH + AGB

A p-value was determined for each factor, indicating the extent to which it was able to explain changes in the response variable. The term with the highest p-value (i.e. least significant) was then removed, and the simplified model was run again; this procedure was repeated until only significant terms were left ($p \leq 0.05$). Table 7.1 shows a summary of the variables tested and the factors found to affect them.

| Response variable | Factors with significant effect |
|---|---|
| No. taxa | none |
| No. unique taxa | none |
| Maturity index | none |
| Diversity (Shannon index) | none |
| Diversity (Simpson index) | treatment (p=0.021); horizontal position (p=0.006); moisture content (p=0.009). |
| Dominance index | treatment (p=0.01); horizontal position (p=0.007); moisture content (p=0.001). |
| Abundance of Tylenchina | vertical position (p=0.007); horizontal position (p=0.017); |
| Abundance of Dorylaimida | sample size (p=0.001); vertical position (p=0.011). |
| Abundance of Enoplida | treatment (p=0.004); vertical position (p=0.004); horizontal position (p=0.02); AGB (p=0.002); |
| Abundance of Mononchida | horizontal position (p=0.008) |

Table 7.1 Results of GLM analysis showing variables tested and factors found to affect them, together with p-values.

Neither the absolute number of taxa nor the number of unique taxa was found to be significantly influenced by any factor tested. The Shannon diversity index showed no significant effects, though horizontal position was on the borderline of significance ($p=0.059$). The Simpson index, on the other hand, was found to be significantly influenced by treatment, horizontal position and moisture content. Figure 7.1 shows a histogram of group means for the Simpson index, indicating that diversity is lowest in the lime plots and highest in the nitrogen+lime plots. The relationship with horizontal position from left to right is negative, i.e. Simpson diversity is highest in the A plots and lowest in the F plots. The relationship with moisture content is also negative – that is, the higher the moisture content, the lower the diversity. The dominance index shows very similar results to the Simpson index. This change is mainly driven by changes in the abundance of the dominant *Helicotylenchus* MOTU – that is, as moisture content increases *Helicotylenchus* becomes more abundant, and other taxa become consequently less abundant, causing measured diversity to decrease.

Of the four major taxonomic groups of nematodes examined, only the enoplids are found to show a significant, direct effect of treatment. As shown in Figure 7.2, they are lowest in the biocide plots and highest in the nitrogen+lime plots. However, enoplid abundance is also found to be significantly affected by spatial position, as are all of the other nematode groups. Dorylaimid abundance is affected by vertical position, but also by sample size.

It is noticeable that most of the response variables tested show a greater effect of spatial position than of experimental treatment, or the environmental parameters measured. This suggests that soil microclimate and spatial heterogeneity are more important than the experimental treatments in determining nematode diversity and taxon distribution.

It is also clear that different nematode groups are responding to different factors, thus any measure of "diversity" *per se* is effectively lumping together a large number of different effects. Therefore it is not necessarily to be expected that a consistent pattern will be found.

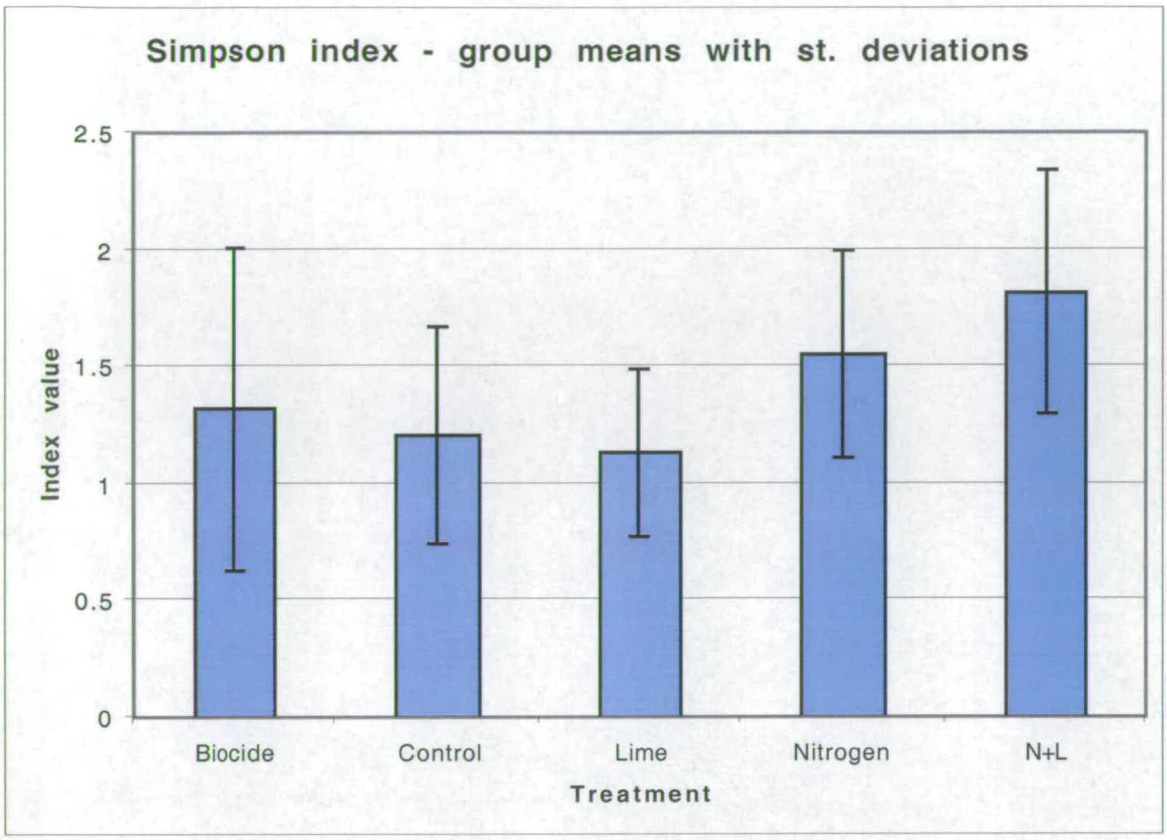



Figure 7.1 Simpson diversity index values (means by treatment groups) for all October samples.

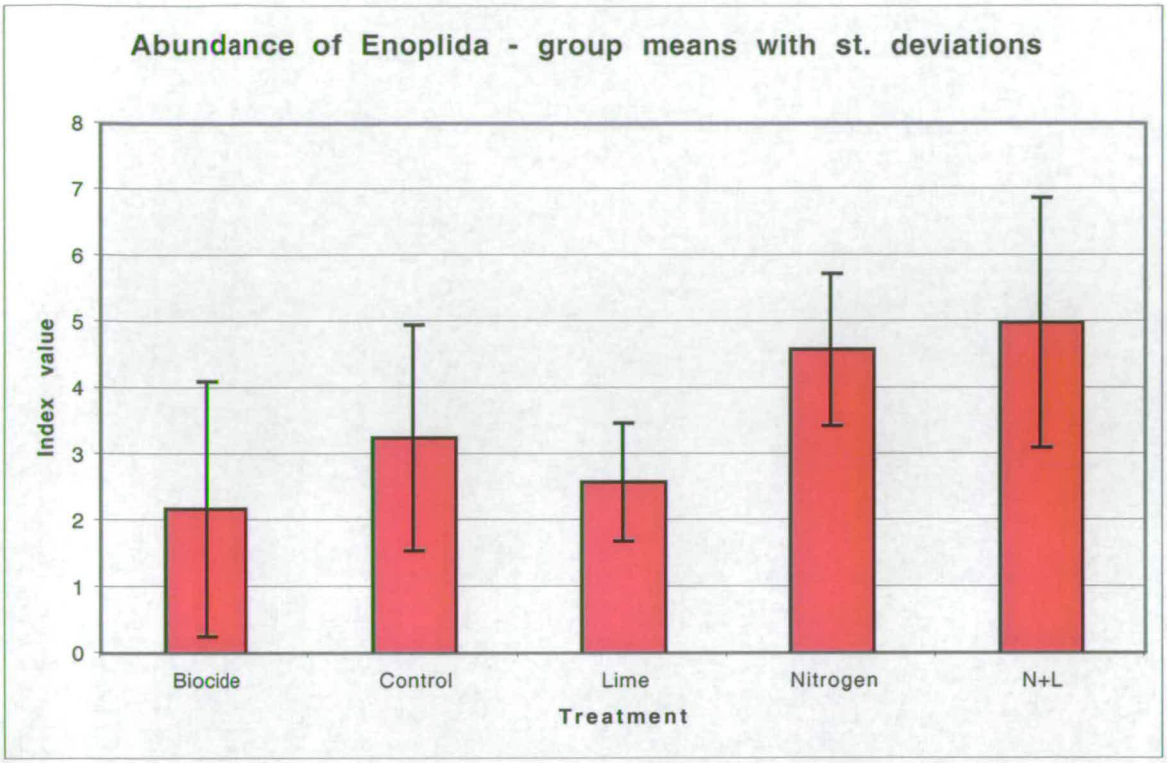



Figure 7.2 Abundances of enoplids (means by treatment groups) for all October samples.

# 8. Discussion and Conclusions

The challenges facing modern taxonomy are not in question (Godfray 2002; Wilson 2003): the diversity of the biosphere is so immense that the number of named and described taxa is dwarfed by those which remain unrecorded, while the number of specialists trained in traditional taxonomic methods is dwindling (De Ley 2000). It is the correct path towards solving these challenges that remains controversial (Lipscomb et al. 2003; Seberg et al. 2003); nevertheless, this work has shown that a DNA-based taxonomic system can provide at least part of a solution to these problems (Blaxter and Floyd 2003).

It has been shown that sequences generated from individual nematodes can be grouped into molecular operational taxonomic units (MOTU). Consequently this allows determination of abundances and patterns of distribution of particular taxa, as well as the calculation of taxon counts and various other indices of diversity, all of which can be correlated with environmental parameters of interest. Additionally, by comparison against a database of sequences from named taxa, MOTU can be assigned to known taxonomic groups, to varying levels of resolution (dictated by the completeness of the reference database). Significantly, however, correlation of MOTU with known taxonomic groups is not in itself a requisite for determining these parameters of diversity: taxon counts can be generated without necessarily knowing anything about the biological attributes of the taxa themselves.

It was found that the number of nematode taxa recorded was very large. In this study, approximately 136 distinct molecular taxa were identified from one site alone. Moreover, the shape of the taxon accumulation curve seen in Figure 4.5.3 (reaching no apparent asymptote) strongly suggests that the nematode community has not been completely sampled in these 2039 individuals. If additional surveys were carried out at the same site, it is likely that the total taxon count would increase considerably: perhaps, if standard statistical estimators are to be believed, to between four and five hundred. When it is considered that only around 200 nematode morphospecies have been described from the British Isles (B. Boag, pers. comm.), these findings would seem to indicate two possibilities. It may be that MOTU do not correspond at all to morphospecies, and that there are a large number of cryptic taxa, distinguishable by SSU sequence but not by morphology; thus, the MOTU approach splits nematodes into more taxa than would be recognised as morphospecies. If, on the other hand, MOTU do approximately correspond to morphospecies this would suggest that British nematodes have been considerably undersampled. These two possiblities are not mutually exclusive, and both may be true to some degree: it is known that the rate of genome evolution varies considerably between difference nematode lineages, and in most cases we have little idea of what level of DNA sequence difference corresponds to what level of morphological distinctness.

In one particular nematode group more information is available. It has been shown, for a set of nematodes in the genus *Panagrolaimus* (cultures isolated from Sourhope soil), that SSU sequence split the five strains into two MOTU, while morphological features were inadequate to discriminate between any of them. Breeding experiments were also carried out, to determine which strains could successfully mate and

produce offspring (a test of the biological species concept, BSC), and it was found that the strains split into two reproductively isolated groups, congruent with the two groups defined by sequence identity (Eyualem and Blaxter 2003). This demonstrates that, at least in this instance, molecular taxon assignments are in agreement with the BSC while morphology is not. However, more research is needed before such conclusions can be judged to apply to nematodes as a whole.

The commonest pattern of taxon abundances observed in nature is the lognormal distribution (Preston 1948), in which a small number of taxa are very common, a small number are very rare, and the majority are of intermediate commonness or rarity. Though the taxon abundances found in this survey did not themselves fit a lognormal distribution (Figure 4.5.2), they are consistent with the hypothesis that the underlying distribution is lognormal: in this survey only the common and intermediate taxa were seen (that is, only the right-hand part of the curve), and the rare taxa would only be encountered if many more individuals were sampled. This would appear to be a strong argument in favour of molecular survey methods, if we are interested in achieving anything approaching exhaustive biodiversity surveys which include rare taxa as well as common ones. Sampling the many thousands of individuals needed to pick up the rarest taxa would be extraordinarily difficult, if not impossible, using morphologcal taxonomy - even for one small site such as Sourhope, let alone the rest of the world. Yet high-throughput sequence-based methods bring such surveys within the realm of the achievable.

We have seen that molecular taxon assignment carries its own particular sources of error and uncertainty. Given that taxa are defined by sequence identity alone, the effect of sequencing errors cannot be ignored; and, given that specimens are placed to taxa by a sequential pairwise clustering algorithm, the effect of sequence processing order must be considered. Problems of this sort are by no means unique to the molecular approach: a morphological taxonomist working through a large number of specimens might, on occasion, mis-measure or erroneously record a certain feature, potentially resulting in a misclassification. And, were a taxonomist to identify the same set of specimens twice over in a randomised processing order, or if the set were identified by two different taxonomists, who can say whether the specimens would be identically classified on each occasion? In the vast majority of cases it is simply not practically possible to carry out such tests, given the global shortage of taxonomists and the numerous pressures on their time. This, therefore, is a significant advantage of molecular taxonomics: it is straightforward to empirically test the magnitude of such sources of error, so that the exact error rate is known. Therefore, as well as an absolute diversity, a range of possible errors can be reported, indicating the degree of uncertainty associated with any diversity measure.

In this survey, it was found that the sequencing error rate, while not insignificant, was low enough that it should have had, at most, a minor influence on the number of taxa recorded. Variation in taxon assignment due to sequence processing order was also appreciable, but did not change the assignment of MOTU to higher taxonomic groups, and in most cases was not sufficient to alter the values of diversity indices; the few samples where diversity index values did change significantly were found to be the result of a few "taxon flocks" of closely related sequences, confined to particular plots – an aspect of the

nematode community that is potentially interesting in itself, brought to our attention by a feature which might otherwise be considered a disadvantage of the molecular approach.

This molecular taxonomic survey method is therefore workable, robust, and capable of measuring both patterns in overall diversity and in specific taxon distributions. However, actual patterns in diversity at the Sourhope study site are complex and difficult to disentangle. It is clear that different nematode taxa are responding differently to particular environmental factors. Certain taxa are influenced by horizontal position, others by vertical position, others by sampling date, and others by the experimental treatments of interest; values such as diversity indices, which combine information about many taxa, are often insensitive to such changes. What is evident is that none of the experimental treatments applied at the Sourhope site – nitrogen, lime or even biocide – has had an effect on overall nematode diversity which is sufficiently strong or unambiguous to be easily distinguished from the existing background "noise" of variation within the nematode community. The spatial heterogeneity of the nematode community is also an unavoidable issue. Nematodes are known to follow patchy distributions, and we cannot necessarily assume that the composition of a single soil core is representative of an entire plot. The ecology of this site is a multidimensional property, as is the diversity of its nematodes and other organisms.

This work demonstrates that the methods are in place for the surveying of nematode communities on a far larger scale: using a standard set of techniques, any scientist with basic training in molecular biology could sample the nematodes of any habitat and produce a set of robust taxonomic assignments. What is more, all such data would be directly comparable between different labs and groups, as DNA sequence data can be easily communicated and software for taxon assignment can be standardised. Within a few years, data of the type produced in this work could be generated from a wide range of ecosystems, providing an unprecedented window into nematode diversity – we could know, for example, which sites are more diverse than others, or have unusual proportions of unique or novel taxa, and which taxa are widely distributed or localised to particular areas.

This is not to suggest that there is no place for morphological information in taxonomy. If we wich to generate information about the taxa themselves, about their biological attributes and their functional roles in natural ecosystems, morphological, ecological and every other kind of information is needed form representatives of those taxa. But this is a separate and distinct task from defining and counting taxa in environmental samples. The advantage of a MOTU approach is that, even where we lack such biological information (as we do for many nematode taxa) this lack of knowledge does not impede us from gaining measures of basic diversity and taxon distributions.

The method of sampling nematodes employed in this study retained the relatively laborious step of manually picking individual nematodes from soil extracts. A more rapid approach, which has been employed in a large number of microbial studies already cited (and also by Markmann 2000 to examine lacustrine meiofauna, including nematodes), is to extract total DNA from a soil, sediment or other environmental sample, PCR all of the SSU (or other gene of interest) copies present, clone the resulting fragments in a library, and sequence the clones. This method has the advantage of high speed and

throughput, but carries the additional uncertainty over how representative the library is of the true genetic diversity, as sequences from certain organisms might show biases in their ease of both PCR and cloning (Anderson et al. 2003). Nevertheless, there is much interest in further developing this type of approach to generate large amounts of diversity data.

As a method, molecular barcoding is likely to be broadly applicable to biodiversity research – virtually all groups of organisms in all habitat types should be amenable to this method of taxonomic analysis. It is therefore unsusprising that increasing numbers of authors are advocating the adoption of a DNA-based taxonomic system (Hebert et al. 2003a; Tautz et al. 2003). In terms of establishing a more general system, perhaps one applicable to all animals, some questions remain, for example the issue of which gene (or set of genes) is appropriate for use as a taxonomic marker. This work has shown that the 5' end of the nuclear SSU is able to successfully delineate taxa within the Nematoda, while Hebert et al. (2003a;b) have argued in favour of the mitochondrial cytochrome oxidase I (COI) as a universal barcode for animal life. To resolve such issues, it would seem necessary to carry out a direct cross comparison of these, and perhaps other genes, by sequencing the genes of interest from the same set of individuals. This would provide a direct test of which markers are most "useful" (by any specified criteria) for molecular taxon assignment.

One issue which has perhaps not been paid sufficient attention by other studies is the technical detail of how taxa are defined from sequences. In prokaryotes, it is common practice to declare that a specified level of similarity in 16S rDNA sequence defines a "species", but different groups have applied this definition inconsistently, with some choosing 99% (Furlong et al. 2002), 98% (Bonnet et al. 2002), or 97% (McCaig et al. 2001). 97% is the closest to a "standard" figure that has been adopted by microbial systematists (Stackebrandt and Goebel 1994; Hägstrom et al. 2002), but the justification for such a practice seems dubious. It has previously been established, based on whole genome DNA-DNA reassociation experiments, that a relative binding ratio of 70% or greater defines conspecific strains (Wayne et al. 1987), and that this value should correspond to an overall sequence identity of around 97%. It is therefore reasoned that a small part of the genome such as the 16S gene should be representative of this level of variation, but there seems no reason to assume that this will be the case, and it has been shown that 16S sequence similarity is not a reliable predictor of DNA-DNA reassociation (Rosselló-Mora and Amann 2001).

Furthermore, these studies do not make it clear how sequences are actually grouped together, whether it is by a sequential pairwise comparison/clustering method such as was employed in this project, or some other approach. Such problems as the effect of sequence processing order and experimental errors in sequencing do not appear to have been taken into account.

The COI system employed by Hebert et al. (2003) was to place specimen sequences to taxa by first creating an alignment of known sequences, and then running a neighbour-joining analysis with one specimen sequence at a time, and assigning the speciment to the same taxonomic group as its nearest neighbour. This appears a problematic approach, as there is no clear definition of how divergent a sequence

must be before it is considered "novel", and hence considerably divergent specimen sequences could be placed in the same taxon because they share the same nearest neigbour. Also the reliability of such an approach is strongly dependent on the set of known taxa chosen to start with: it is necessary to "know" in advance which taxa you are looking for, and different starting datasets could conceivable give entirely different taxon assigments for the same set of sequences.

The approach of Hebert et al. appears to retain something of an "idealistic" species concept: when a set of unknown specimens is identified, each one is compared to some external set of taxonomic identifications, and classified relative to these knowns. In contrast, the approach taken in this work is to classify a set of specimens only relative to each other. What is defined before running the process is not a set of taxa which are expected to be found, but only a set of rules governing how to define and name taxa. The relative "correctness" of each approach will doubtless be subject to much debate, but it might be argued that our MOTU approach carries with it the least philosophical "baggage" in terms of unstated assumptions and preconcieved ideas about what taxa exist. The MOTU definition process can be run with absolutely no *a priori* assumptions about what taxa already exist or what is expected to be found: the specimens in a sample are clustered only according to the information present in that sample.

Of course, the nature of the process means that it would be straightforward to modify it so that taxon assignments are made relative to an existing set of taxa, if so desired. This could be achieved by first creating a set of sequences belonging to known taxa, and modifying the program so that new sequences, instead of being searched against each other, are searched against this set of named sequences, and assigned to the same taxon if they are sufficiently similar, or given some name such as "novel01", "novel02" etc. if they are not.

A potential difficulty of any molecular barcoding system is the problem of encountering unknown taxa – that is, if a MOTU sequence discovered in a survey cannot be matched to any known sequence, we are unable to say what it is. The situation will, however, only improve as more named taxa are added to databases for comparison. Indeed, an equivalent problem exists in morphological taxonomy – in any survey there is a possibility that a new specimen does not match any previously described morphospecies (when a novel environment is sampled such specimens may even consitute a majority of the survey (Lawton et al. 1996). If anything, a molecular system makes it easier to deal with such taxa: any novel sequence can be considered simply as another taxon which counts toward the overall diversity, with a defined degree of genetic distance from its nearest relative, and no further description is needed. When a novel morphospecies is encountered, a taxonomist first faces a laborious search through the literature to be certain that it is indeed novel; and, having determined that it is so, must subsequently produce and publish a morphological description before any other taxonomist is able to make use of this information. Indeed, the study of Lawton et al. did not attempt this kind of description but simply placed unidentifiable individuals to "morphospecies" based on the information available to obtain taxon counts, meaning that this information cannot be directly compared with other surveys. The MOTU approach, in contrast, allows both straightforward taxon assignment and comparison between surveys.

It is possible that, for groups of organisms such as nematodes, where the number of individuals and taxa is simply too great for the traditional taxonomic process, a molecular system may well be the *only* realistic way to catalogue the diversity that exists, to anything approaching a complete "taxon list", within a human lifetime.

# Appendix 1: List of Sequences

This section contains a list of all 2039 single-nematode SSU sequences generated in this project, together with the MOTU to which they belong (in the standard 2bp_MOTU set shown in Figure 4.7.1), the Sourhope plots from which they were sampled, and the data of sampling. An "unknown" plot origin means that the sequence is part of the preliminary survey in which nematodes were extracted from a unknown soil sample, and thus the actual plot of origin is not known.

Also indicated here is the worker responsible for producing the sequence: who picked the nematode from the soil extract, who carried out the PCR from the nematode, and sequenced the PCR product (RF=Robin Floyd; EA=Eyualem Abebe; MW=Mark Welsh).

The second table lists all sequences from other studies which were used in the phylogenetic analysis, with their GenBank accession numbers (for submitted sequences) or the source (for unpublished data).

Table 1: sequences generated in this study.

| Sequence name | Plot origin | Sample date | MOTU | Pick | PCR | Sequencing |
|---|---|---|---|---|---|---|
| 10102ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10105ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10108ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10110ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10111ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10112ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10113ED.seq | unknown | 27/7/00 | 2bp_MOTU0009 | RF | RF | RF |
| 10114ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10115ED.seq | unknown | 27/7/00 | 2bp_MOTU0035 | RF | RF | RF |
| 10116ED.seq | unknown | 27/7/00 | 2bp_MOTU0129 | RF | RF | RF |
| 10117ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10120ED.seq | unknown | 27/7/00 | 2bp_MOTU0033 | RF | RF | RF |
| 10121ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10122ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10123ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10124ED.seq | unknown | 27/7/00 | 2bp_MOTU0052 | RF | RF | RF |
| 10125ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10126ED.seq | unknown | 27/7/00 | 2bp_MOTU0019 | RF | RF | RF |
| 10130ED.seq | unknown | 27/7/00 | 2bp_MOTU0019 | RF | RF | RF |
| 10132ED.seq | unknown | 27/7/00 | 2bp_MOTU0019 | RF | RF | RF |
| 10133ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10134ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10137ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10138ED.seq | unknown | 27/7/00 | 2bp_MOTU0025 | RF | RF | RF |
| 10139ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10141ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10145ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10148ED.seq | unknown | 27/7/00 | 2bp_MOTU0019 | RF | RF | RF |
| 10149ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10151ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10152ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10153ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10154ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10155ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10156ED.seq | unknown | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10157ED.seq | unknown | 27/7/00 | 2bp_MOTU0080 | RF | RF | RF |
| 10162ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10163ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10164ED.seq | unknown | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10165ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10166ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10168ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10169ED.seq | unknown | 27/7/00 | 2bp_MOTU0009 | RF | RF | RF |
| 10170ED.seq | unknown | 27/7/00 | 2bp_MOTU0090 | RF | RF | RF |
| 10173ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10174ED.seq | unknown | 27/7/00 | 2bp_MOTU0089 | RF | RF | RF |
| 10175ED.seq | unknown | 27/7/00 | 2bp_MOTU0004 | RF | RF | RF |
| 10180ED.seq | unknown | 27/7/00 | 2bp_MOTU0091 | RF | RF | RF |
| 10181ED.seq | unknown | 27/7/00 | 2bp_MOTU0016 | RF | RF | RF |
| 10182ED.seq | unknown | 27/7/00 | 2bp_MOTU0033 | RF | RF | RF |
| 10185ED.seq | unknown | 27/7/00 | 2bp_MOTU0084 | RF | RF | RF |
| 10188ED.seq | unknown | 27/7/00 | 2bp_MOTU0009 | RF | RF | RF |
| 10190ED.seq | unknown | 27/7/00 | 2bp_MOTU0091 | RF | RF | RF |
| 10191ED.seq | unknown | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10193ED.seq | unknown | 27/7/00 | 2bp_MOTU0009 | RF | RF | RF |
| 10195ED.seq | unknown | 27/7/00 | 2bp_MOTU0009 | RF | RF | RF |
| 10202ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10203ED.seq | 4D | 27/7/00 | 2bp_MOTU0033 | RF | RF | RF |
| 10205ED.seq | 4D | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10206ED.seq | 4D | 27/7/00 | 2bp_MOTU0039 | RF | RF | RF |
| 10207ED.seq | 4D | 27/7/00 | 2bp_MOTU0006 | RF | RF | RF |
| 10208ED.seq | 4D | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10209ED.seq | 4D | 27/7/00 | 2bp_MOTU0033 | RF | RF | RF |
| 10210ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10211ED.seq | 4D | 27/7/00 | 2bp_MOTU0119 | RF | RF | RF |
| 10217ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10218ED.seq | 4D | 27/7/00 | 2bp_MOTU0013 | RF | RF | RF |
| 10219ED.seq | 4D | 27/7/00 | 2bp_MOTU0002 | RF | RF | RF |
| 10222ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10224ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10227ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10229ED.seq | 4D | 27/7/00 | 2bp_MOTU0037 | RF | RF | RF |
| 10230ED.seq | 4D | 27/7/00 | 2bp_MOTU0053 | RF | RF | RF |
| 10231ED.seq | 4D | 27/7/00 | 2bp_MOTU0017 | RF | RF | RF |
| 10232ED.seq | 4D | 27/7/00 | 2bp_MOTU0040 | RF | RF | RF |
| 10405ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | RF | RF | RF |
| 10406ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | RF | RF | RF |
| 10407ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 10408ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 10413ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0082 | RF | RF | RF |
| 10414ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0117 | RF | RF | RF |
| 10417ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0067 | RF | RF | RF |
| 10419ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0064 | RF | RF | RF |
| 10420ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0060 | RF | RF | RF |
| 10421ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0044 | RF | RF | RF |
| 10502ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 10506ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 10507ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0127 | MW | RF | RF |
| 10508ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | MW | RF | RF |
| 10509ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0060 | MW | RF | RF |
| 10511ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0010 | MW | RF | RF |
| 10512ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0030 | MW | RF | RF |
| 10513ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0009 | MW | RF | RF |
| 10514ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 10515ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 10517ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0093 | MW | RF | RF |
| 10518ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | MW | RF | RF |
| 10519ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0007 | MW | RF | RF |
| 10520ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | MW | RF | RF |
| 10521ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | MW | RF | RF |
| 10522ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | MW | RF | RF |
| 10601ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 10604ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10606ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10607ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 10609ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10610ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0107 | EA | RF | RF |
| 10613ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0060 | EA | RF | RF |
| 10616ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 10618ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0098 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10620ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 10701ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | EA | RF | RF |
| 10702ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | EA | RF | RF |
| 10705ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0011 | EA | RF | RF |
| 10706ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 10707ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 10708ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 10709ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 10710ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10711ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 10712ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10713ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0010 | EA | RF | RF |
| 10714ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 10715ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0010 | EA | RF | RF |
| 10719ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0031 | EA | RF | RF |
| 10724ED.seq | 1F_upper | 12/6/01 | 2bp_MOTU0020 | EA | RF | RF |
| 10801ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10802ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10803ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0049 | RF | EA | RF |
| 10805ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0078 | RF | EA | RF |
| 10806ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10807ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10810ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10811ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10813ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10814ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10815ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10816ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10817ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10818ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10819ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10820ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10821ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10822ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 10901ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 10903ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10904ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10905ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 10906ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 10907ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 10909ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10911ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 10912ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10913ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 10914ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0020 | RF | RF | RF |
| 10916ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0020 | RF | RF | RF |
| 10917ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 10918ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0020 | RF | RF | RF |
| 10919ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0025 | RF | RF | RF |
| 10920ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0058 | RF | RF | RF |
| 10921ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0101 | RF | RF | RF |
| 10922ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10923ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 10924ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11101ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0057 | RF | EA | RF |
| 11102ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11103ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11106ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11110ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11111ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11113ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0010 | RF | EA | RF |
| 11115ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11116ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11118ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11123ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11124ED.seq | 1F_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11223ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11302ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0128 | RF | EA | RF |
| 11303ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11304ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11305ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11306ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11310ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 11312ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11313ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11314ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0012 | RF | EA | RF |
| 11401ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0026 | EA | RF | RF |
| 11402ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11403ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 11404ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 11405ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0031 | EA | RF | RF |
| 11406ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0020 | EA | RF | RF |
| 11407ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11408ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 11411ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11413ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11414ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11417ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0004 | EA | RF | RF |
| 11419ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11420ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11421ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0030 | EA | RF | RF |
| 11422ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11502ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11504ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11505ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11506ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11507ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11508ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11509ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11510ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11511ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11512ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11514ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11515ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11516ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11517ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0030 | RF | EA | RF |
| 11518ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0031 | RF | EA | RF |
| 11519ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11520ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11521ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11522ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11523ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11524ED.seq | 2B_upper | 12/6/01 | 2bp_MOTU0030 | RF | EA | RF |
| 11601ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 11603ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | EA | RF | RF |
| 11604ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0140 | EA | RF | RF |
| 11607ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11608ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | EA | RF | RF |
| 11610ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11611ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11612ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 11614ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11616ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11617ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11618ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11619ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | EA | RF | RF |
| 11620ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11621ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11622ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | EA | RF | RF |
| 11623ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | EA | RF | RF |
| 11624ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 11701ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11704ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0026 | RF | EA | RF |
| 11707ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11708ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11710ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11714ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 11715ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0054 | RF | EA | RF |
| 11718ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 11719ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11723ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 11802ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11803ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11804ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11807ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11809ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 11811ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11814ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11816ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11817ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0012 | RF | RF | RF |
| 11819ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 11821ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11822ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 11824ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | RF | RF | RF |
| 11902ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | MW | RF | RF |
| 11903ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | MW | RF | RF |
| 11908ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | MW | RF | RF |
| 11916ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 11918ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0006 | MW | RF | RF |
| 11920ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | MW | RF | RF |
| 11921ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0125 | MW | RF | RF |
| 11923ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0018 | MW | RF | RF |
| 11924ED.seq | 2B_lower | 12/6/01 | 2bp_MOTU0002 | MW | RF | RF |
| 12002ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0076 | RF | RF | RF |
| 12003ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12004ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12005ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0011 | RF | RF | RF |
| 12007ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12008ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12009ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0010 | RF | RF | RF |
| 12010ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12011ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | RF | RF | RF |
| 12012ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 12013ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12014ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 12015ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12016ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | RF | RF | RF |
| 12017ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12018ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12019ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 12020ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | RF |
| 12021ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 12022ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | RF | RF |
| 12023ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0026 | RF | RF | RF |
| 12101ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12102ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12103ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12104ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12105ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12106ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12107ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0030 | RF | EA | RF |
| 12110ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 12111ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12112ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12113ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0068 | RF | EA | RF |
| 12114ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12115ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 12116ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12117ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12118ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12120ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 12122ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12123ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 12124ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12201ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12204ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | RF | EA | RF |
| 12207ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0016 | RF | EA | RF |
| 12209ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12211ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 12213ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0008 | RF | EA | RF |
| 12214ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 12217ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0016 | RF | EA | RF |
| 12218ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0012 | RF | EA | RF |
| 12223ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | RF | EA | RF |
| 12301ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12302ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12303ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12304ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12305ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 12306ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12307ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0026 | EA | RF | RF |
| 12309ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0026 | EA | RF | RF |
| 12310ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 12311ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0006 | EA | RF | RF |
| 12312ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 12313ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12314ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12315ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12317ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0010 | EA | RF | RF |
| 12318ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0103 | EA | RF | RF |
| 12319ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 12320ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12321ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12322ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12323ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12324ED.seq | 3D_upper | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 12401ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12402ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | EA |
| 12406ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12407ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | EA |
| 12408ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12409ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0010 | RF | EA | EA |
| 12410ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12411ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12414ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12415ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12416ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0018 | RF | EA | EA |
| 12418ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12419ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12420ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12422ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12423ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12424ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12523ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0011 | MW | EA | EA |
| 12601ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0012 | EA | EA | EA |
| 12602ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0012 | EA | EA | EA |
| 12603ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12604ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 12605ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0012 | EA | EA | EA |
| 12607ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 12608ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12609ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12612ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0016 | EA | EA | EA |
| 12613ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12614ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12615ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12616ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12617ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12620ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0012 | EA | EA | EA |
| 12622ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 12623ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 12624ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12702ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0060 | EA | EA | EA |
| 12707ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0098 | EA | EA | EA |
| 12709ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12711ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12712ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12713ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12714ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12715ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 12716ED.seq | 3D_lower | 12/6/01 | 2bp_MOTU0010 | EA | EA | EA |
| 12802ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12803ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12804ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12805ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12806ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12807ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12808ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0007 | RF | EA | EA |
| 12809ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0009 | RF | EA | EA |
| 12813ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12815ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0026 | RF | EA | EA |
| 12817ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |

| | | | | | | |
|---|---|---|---|---|---|---|
| 12819ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12821ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12823ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12824ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | EA |
| 12901ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12902ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12903ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12904ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12905ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12906ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12907ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12908ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0006 | RF | RF | EA |
| 12909ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0006 | RF | RF | EA |
| 12910ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12913ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12914ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12915ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12916ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12918ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12920ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0010 | RF | RF | EA |
| 12921ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0018 | RF | RF | EA |
| 12922ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12923ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 12924ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | RF | RF | EA |
| 13003ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13004ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13005ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13006ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13007ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13008ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0033 | EA | EA | EA |
| 13009ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0114 | EA | EA | EA |
| 13010ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13011ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0033 | EA | EA | EA |
| 13012ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13013ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | EA |
| 13016ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0031 | EA | EA | EA |
| 13018ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0029 | EA | EA | EA |
| 13019ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0033 | EA | EA | EA |
| 13020ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13021ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 13022ED.seq | 4D_upper | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 13201ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13202ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13203ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13204ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13205ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13206ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13207ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13208ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13209ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13210ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13211ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13212ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13213ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13214ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13215ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13216ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13217ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13218ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13220ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13221ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13222ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13223ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 13224ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13301ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13302ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13303ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13305ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13306ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13307ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13308ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13309ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13310ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 13313ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13314ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13315ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13316ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13317ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13318ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13319ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13320ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13321ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13322ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13323ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13324ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13401ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | RF |
| 13405ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | RF |
| 13409ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | RF |
| 13412ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | RF |
| 13418ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0022 | EA | EA | RF |
| 13419ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0033 | EA | EA | RF |
| 13424ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | RF |
| 13502ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13503ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13504ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13506ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13507ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13508ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13509ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13511ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13513ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 13514ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13515ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13516ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13517ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0018 | EA | RF | RF |
| 13518ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13519ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13520ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13521ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13522ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13523ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0002 | EA | RF | RF |
| 13524ED.seq | 4D_lower | 12/6/01 | 2bp_MOTU0049 | EA | RF | RF |
| 13601ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13602ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13603ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13604ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0025 | RF | EA | RF |
| 13605ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0016 | RF | EA | RF |
| 13606ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13607ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13609ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0031 | RF | EA | RF |
| 13610ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 13611ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0108 | RF | EA | RF |
| 13612ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0031 | RF | EA | RF |
| 13615ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0009 | RF | EA | RF |
| 13617ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0030 | RF | EA | RF |
| 13618ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | RF | EA | RF |
| 13619ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0009 | RF | EA | RF |
| 13622ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0029 | RF | EA | RF |
| 13702ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | EA | EA | RF |
| 13703ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0010 | EA | EA | RF |
| 13706ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13708ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13711ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0018 | EA | EA | RF |
| 13713ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0006 | EA | EA | RF |
| 13717ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13719ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13720ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13723ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13724ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | EA | EA | RF |
| 13805ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0020 | EA | EA | RF |
| 13902ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 13903ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13904ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0011 | RF | EA | RF |
| 13905ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0011 | RF | EA | RF |
| 13906ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13907ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 13908ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13909ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0031 | RF | EA | RF |
| 13911ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0010 | RF | EA | RF |
| 13913ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0026 | RF | EA | RF |
| 13915ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0002 | RF | EA | RF |
| 13916ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0049 | RF | EA | RF |
| 13917ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0026 | RF | EA | RF |
| 13920ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0004 | RF | EA | RF |
| 13922ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0049 | RF | EA | RF |
| 13924ED.seq | 5A_upper | 12/6/01 | 2bp_MOTU0011 | RF | EA | RF |
| 14201ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14202ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14203ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14206ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14208ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14209ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14210ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14212ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14215ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14216ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14217ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14218ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14219ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14220ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14224ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14301ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14302ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 14303ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 14307ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 14308ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0006 | EA | EA | EA |
| 14309ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0018 | EA | EA | EA |
| 14318ED.seq | 5A_lower | 12/6/01 | 2bp_MOTU0002 | EA | EA | EA |
| 15003ED.seq | 1A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15004ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15005ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15006ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15007ED.seq | 1A | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15008ED.seq | 1A | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15010ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15011ED.seq | 1A | 22/10/01 | 2bp_MOTU0048 | RF | RF | RF |
| 15012ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15013ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15014ED.seq | 1A | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 15018ED.seq | 1A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15020ED.seq | 1A | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 15022ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15023ED.seq | 1A | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15024ED.seq | 1A | 22/10/01 | 2bp_MOTU0087 | RF | RF | RF |
| 15025ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15027ED.seq | 1A | 22/10/01 | 2bp_MOTU0093 | RF | RF | RF |
| 15028ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15029ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15030ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15031ED.seq | 1A | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 15032ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15033ED.seq | 1A | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 15034ED.seq | 1A | 22/10/01 | 2bp_MOTU0135 | RF | RF | RF |
| 15035ED.seq | 1A | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15036ED.seq | 1A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15037ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15039ED.seq | 1A | 22/10/01 | 2bp_MOTU0069 | RF | RF | RF |
| 15040ED.seq | 1A | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15041ED.seq | 1A | 22/10/01 | 2bp_MOTU0120 | RF | RF | RF |
| 15042ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15043ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15044ED.seq | 1A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15045ED.seq | 1A | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15046ED.seq | 1A | 22/10/01 | 2bp_MOTU0034 | RF | RF | RF |
| 15049ED.seq | 1A | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 15050ED.seq | 1A | 22/10/01 | 2bp_MOTU0037 | RF | RF | RF |
| 15051ED.seq | 1A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15052ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15053ED.seq | 1A | 22/10/01 | 2bp_MOTU0048 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15054ED.seq | 1A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15055ED.seq | 1A | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 15056ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15057ED.seq | 1A | 22/10/01 | 2bp_MOTU0014 | RF | RF | RF |
| 15058ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15059ED.seq | 1A | 22/10/01 | 2bp_MOTU0110 | RF | RF | RF |
| 15061ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15062ED.seq | 1A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15063ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15064ED.seq | 1A | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15065ED.seq | 1A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15066ED.seq | 1A | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 15067ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15068ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15070ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15071ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15072ED.seq | 1A | 22/10/01 | 2bp_MOTU0105 | RF | RF | RF |
| 15073ED.seq | 1A | 22/10/01 | 2bp_MOTU0037 | RF | RF | RF |
| 15074ED.seq | 1A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15075ED.seq | 1A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15076ED.seq | 1A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15077ED.seq | 1A | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 15078ED.seq | 1A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15079ED.seq | 1A | 22/10/01 | 2bp_MOTU0085 | RF | RF | RF |
| 15080ED.seq | 1A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 15081ED.seq | 1A | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15082ED.seq | 1A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15101ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15102ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15105ED.seq | 1B | 22/10/01 | 2bp_MOTU0021 | RF | RF | RF |
| 15106ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15107ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15109ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15110ED.seq | 1B | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 15111ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15112ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15113ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15114ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15115ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15117ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15118ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15119ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15120ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15121ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15122ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15123ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15124ED.seq | 1B | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15125ED.seq | 1B | 22/10/01 | 2bp_MOTU0040 | RF | RF | RF |
| 15126ED.seq | 1B | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 15128ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15129ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15130ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15132ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15133ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15134ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15135ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15136ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15137ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15138ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15139ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15140ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15141ED.seq | 1B | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15142ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15143ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15144ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15145ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15146ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15147ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15149ED.seq | 1B | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 15150ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15151ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15152ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15153ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15154ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15155ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15157ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15158ED.seq | 1B | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15159ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15160ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15161ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15162ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15164ED.seq | 1B | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15165ED.seq | 1B | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15166ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15168ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15169ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15170ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15171ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15172ED.seq | 1B | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15173ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15174ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15175ED.seq | 1B | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15176ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15177ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15178ED.seq | 1B | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15179ED.seq | 1B | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15203ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15204ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15205ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15206ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15208ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15210ED.seq | 1C | 22/10/01 | 2bp_MOTU0048 | EA | RF | RF |
| 15211ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15212ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15214ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15215ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15216ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15217ED.seq | 1C | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 15218ED.seq | 1C | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 15219ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15220ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15221ED.seq | 1C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15222ED.seq | 1C | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 15223ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15224ED.seq | 1C | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 15225ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15226ED.seq | 1C | 22/10/01 | 2bp_MOTU0010 | EA | RF | RF |
| 15227ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15228ED.seq | 1C | 22/10/01 | 2bp_MOTU0058 | EA | RF | RF |
| 15229ED.seq | 1C | 22/10/01 | 2bp_MOTU0048 | EA | RF | RF |
| 15230ED.seq | 1C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15231ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15232ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15233ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15234ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15236ED.seq | 1C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15238ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15239ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15240ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15241ED.seq | 1C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15242ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15243ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15244ED.seq | 1C | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 15245ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15246ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15247ED.seq | 1C | 22/10/01 | 2bp_MOTU0124 | EA | RF | RF |
| 15248ED.seq | 1C | 22/10/01 | 2bp_MOTU0029 | EA | RF | RF |
| 15249ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15250ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15251ED.seq | 1C | 22/10/01 | 2bp_MOTU0058 | EA | RF | RF |
| 15253ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15255ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15256ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15257ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15258ED.seq | 1C | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15259ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15260ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15261ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15265ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15267ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15268ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15269ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15270ED.seq | 1C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15271ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15274ED.seq | 1C | 22/10/01 | 2bp_MOTU0025 | EA | RF | RF |
| 15276ED.seq | 1C | 22/10/01 | 2bp_MOTU0009 | EA | RF | RF |
| 15277ED.seq | 1C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15278ED.seq | 1C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15281ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15282ED.seq | 1C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15302ED.seq | 1E | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15303ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15304ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15305ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15308ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15313ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15314ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15315ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15316ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15318ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15319ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15321ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15322ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15323ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15324ED.seq | 1E | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15326ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15333ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15334ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15335ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15336ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15337ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15340ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15342ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15343ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15344ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15345ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15346ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15348ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15349ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15350ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15352ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15353ED.seq | 1E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15354ED.seq | 1E | 22/10/01 | 2bp_MOTU0099 | RF | RF | RF |
| 15355ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15356ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15357ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15358ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15359ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15360ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15362ED.seq | 1E | 22/10/01 | 2bp_MOTU0072 | RF | RF | RF |
| 15363ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15364ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15365ED.seq | 1E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15367ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15368ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15369ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15370ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15371ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15372ED.seq | 1E | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 15373ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15374ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15375ED.seq | 1E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15376ED.seq | 1E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15378ED.seq | 1E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15402ED.seq | 1F | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15403ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15405ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15406ED.seq | 1F | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15407ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15408ED.seq | 1F | 22/10/01 | 2bp_MOTU0097 | RF | RF | RF |
| 15409ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15410ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15411ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15412ED.seq | 1F | 22/10/01 | 2bp_MOTU0131 | RF | RF | RF |
| 15413ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15414ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15415ED.seq | 1F | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15416ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15417ED.seq | 1F | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15418ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15419ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15420ED.seq | 1F | 22/10/01 | 2bp_MOTU0009 | RF | RF | RF |
| 15421ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15425ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15426ED.seq | 1F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15427ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15428ED.seq | 1F | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15429ED.seq | 1F | 22/10/01 | 2bp_MOTU0109 | RF | RF | RF |
| 15430ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15431ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15432ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15433ED.seq | 1F | 22/10/01 | 2bp_MOTU0113 | RF | RF | RF |
| 15434ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15435ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15436ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15437ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15438ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15439ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15440ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15442ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15443ED.seq | 1F | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15444ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15445ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15446ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15447ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15448ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15449ED.seq | 1F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15451ED.seq | 1F | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15452ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15453ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15454ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15455ED.seq | 1F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15456ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15458ED.seq | 1F | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 15459ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15460ED.seq | 1F | 22/10/01 | 2bp_MOTU0061 | RF | RF | RF |
| 15461ED.seq | 1F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15463ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15464ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15465ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15466ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15467ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15468ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15469ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15470ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15471ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15472ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15474ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15475ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15476ED.seq | 1F | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15477ED.seq | 1F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15478ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15479ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15480ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15482ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15483ED.seq | 1F | 22/10/01 | 2bp_MOTU0097 | RF | RF | RF |
| 15484ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15485ED.seq | 1F | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 15486ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15487ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15488ED.seq | 1F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15489ED.seq | 1F | 22/10/01 | 2bp_MOTU0029 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15490ED.seq | 1F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15501ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15502ED.seq | 2A | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 15503ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15504ED.seq | 2A | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15505ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15506ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15507ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15509ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15510ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15511ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15512ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15513ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15514ED.seq | 2A | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15515ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15516ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15517ED.seq | 2A | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15518ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15519ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15521ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15522ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15523ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15524ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15525ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15526ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15527ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15528ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15529ED.seq | 2A | 22/10/01 | 2bp_MOTU0012 | EA | RF | RF |
| 15530ED.seq | 2A | 22/10/01 | 2bp_MOTU0043 | EA | RF | RF |
| 15531ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15532ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15533ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15534ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15535ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15536ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15537ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15538ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15541ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15542ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15543ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15544ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15545ED.seq | 2A | 22/10/01 | 2bp_MOTU0048 | EA | RF | RF |
| 15546ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15547ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15548ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15549ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15550ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15551ED.seq | 2A | 22/10/01 | 2bp_MOTU0078 | EA | RF | RF |
| 15552ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15553ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15554ED.seq | 2A | 22/10/01 | 2bp_MOTU0086 | EA | RF | RF |
| 15555ED.seq | 2A | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15556ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15557ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15558ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15560ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15566ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15567ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15568ED.seq | 2A | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15569ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15570ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15571ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15572ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15573ED.seq | 2A | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15574ED.seq | 2A | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15575ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15576ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15577ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15578ED.seq | 2A | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 15579ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15580ED.seq | 2A | 22/10/01 | 2bp_MOTU0032 | EA | RF | RF |
| 15581ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |
| 15582ED.seq | 2A | 22/10/01 | 2bp_MOTU0027 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15587ED.seq | 2A | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15601ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15603ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15604ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15605ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15606ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15607ED.seq | 2B | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 15612ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15613ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15614ED.seq | 2B | 22/10/01 | 2bp_MOTU0048 | EA | RF | RF |
| 15615ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15616ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15620ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15621ED.seq | 2B | 22/10/01 | 2bp_MOTU0030 | EA | RF | RF |
| 15622ED.seq | 2B | 22/10/01 | 2bp_MOTU0010 | EA | RF | RF |
| 15623ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15624ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15625ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15626ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15633ED.seq | 2B | 22/10/01 | 2bp_MOTU0137 | EA | RF | RF |
| 15634ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15635ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15637ED.seq | 2B | 22/10/01 | 2bp_MOTU0031 | EA | RF | RF |
| 15639ED.seq | 2B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 15641ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15642ED.seq | 2B | 22/10/01 | 2bp_MOTU0125 | EA | RF | RF |
| 15643ED.seq | 2B | 22/10/01 | 2bp_MOTU0102 | EA | RF | RF |
| 15647ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15648ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15649ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15650ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15651ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15652ED.seq | 2B | 22/10/01 | 2bp_MOTU0060 | EA | RF | RF |
| 15653ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15654ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15656ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15657ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15658ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15660ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15661ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15662ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15663ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15664ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15665ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15666ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15667ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15669ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15670ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15671ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15672ED.seq | 2B | 22/10/01 | 2bp_MOTU0109 | EA | RF | RF |
| 15674ED.seq | 2B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 15675ED.seq | 2B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 15676ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15677ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15678ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15679ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15681ED.seq | 2B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 15683ED.seq | 2B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 15702ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15703ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15705ED.seq | 2C | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15707ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15708ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15709ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15710ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15711ED.seq | 2C | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 15712ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15713ED.seq | 2C | 22/10/01 | 2bp_MOTU0096 | RF | RF | RF |
| 15714ED.seq | 2C | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15715ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15716ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15717ED.seq | 2C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15718ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15720ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15721ED.seq | 2C | 22/10/01 | 2bp_MOTU0023 | RF | RF | RF |
| 15722ED.seq | 2C | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15723ED.seq | 2C | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15725ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15726ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15727ED.seq | 2C | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15730ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15732ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15734ED.seq | 2C | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15735ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15737ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15740ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15741ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15744ED.seq | 2C | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15747ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15749ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15750ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15751ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15752ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15753ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15755ED.seq | 2C | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15756ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15757ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15762ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15763ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15764ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15775ED.seq | 2C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15776ED.seq | 2C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15777ED.seq | 2C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15783ED.seq | 2C | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15802ED.seq | 2D | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15804ED.seq | 2D | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 15805ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15806ED.seq | 2D | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 15807ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15808ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15812ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15815ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15817ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15822ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15825ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15827ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15833ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15837ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15840ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15846ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15847ED.seq | 2D | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 15854ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15855ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15860ED.seq | 2D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15861ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15862ED.seq | 2D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15869ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15870ED.seq | 2D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 15871ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15872ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15873ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15874ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15875ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15877ED.seq | 2D | 22/10/01 | 2bp_MOTU0017 | RF | RF | RF |
| 15878ED.seq | 2D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15879ED.seq | 2D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15880ED.seq | 2D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15881ED.seq | 2D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15882ED.seq | 2D | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 15883ED.seq | 2D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15902ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15904ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15905ED.seq | 2E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 15906ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15907ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15908ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15909ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15910ED.seq | 2E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15911ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15912ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15913ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15914ED.seq | 2E | 22/10/01 | 2bp_MOTU0034 | RF | RF | RF |
| 15915ED.seq | 2E | 22/10/01 | 2bp_MOTU0126 | RF | RF | RF |
| 15916ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15918ED.seq | 2E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15919ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15920ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15921ED.seq | 2E | 22/10/01 | 2bp_MOTU0062 | RF | RF | RF |
| 15922ED.seq | 2E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15923ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15924ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15925ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15926ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15927ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15928ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15929ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15930ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15931ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15932ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15933ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15934ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15935ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15936ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15937ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15938ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15939ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15940ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15942ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15943ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15944ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15945ED.seq | 2E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 15946ED.seq | 2E | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 15947ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15948ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15949ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15950ED.seq | 2E | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15951ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15952ED.seq | 2E | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 15953ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15954ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15955ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15956ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15957ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15958ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15959ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15960ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 15961ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15962ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15963ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15964ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15965ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15966ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15967ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15968ED.seq | 2E | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 15969ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15970ED.seq | 2E | 22/10/01 | 2bp_MOTU0070 | RF | RF | RF |
| 15971ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15972ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15973ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15974ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15975ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15976ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15977ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15978ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15979ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15980ED.seq | 2E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 15981ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 15982ED.seq | 2E | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 15983ED.seq | 2E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15984ED.seq | 2E | 22/10/01 | 2bp_MOTU0100 | RF | RF | RF |
| 15985ED.seq | 2E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16002ED.seq | 3A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16007ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16008ED.seq | 3A | 22/10/01 | 2bp_MOTU0122 | RF | RF | RF |
| 16009ED.seq | 3A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16010ED.seq | 3A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16012ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16013ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16014ED.seq | 3A | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 16015ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16016ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16017ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16018ED.seq | 3A | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 16019ED.seq | 3A | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 16020ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16021ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16022ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16023ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16024ED.seq | 3A | 22/10/01 | 2bp_MOTU0045 | RF | RF | RF |
| 16026ED.seq | 3A | 22/10/01 | 2bp_MOTU0034 | RF | RF | RF |
| 16027ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16028ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16029ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16030ED.seq | 3A | 22/10/01 | 2bp_MOTU0048 | RF | RF | RF |
| 16031ED.seq | 3A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16032ED.seq | 3A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16033ED.seq | 3A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16034ED.seq | 3A | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16035ED.seq | 3A | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16036ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16037ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16038ED.seq | 3A | 22/10/01 | 2bp_MOTU0009 | RF | RF | RF |
| 16039ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16040ED.seq | 3A | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 16041ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16042ED.seq | 3A | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 16043ED.seq | 3A | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16046ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16047ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16048ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16049ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16050ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16051ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16052ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16053ED.seq | 3A | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 16054ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16055ED.seq | 3A | 22/10/01 | 2bp_MOTU0048 | RF | RF | RF |
| 16056ED.seq | 3A | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 16057ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16065ED.seq | 3A | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 16066ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16067ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16068ED.seq | 3A | 22/10/01 | 2bp_MOTU0034 | RF | RF | RF |
| 16069ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16070ED.seq | 3A | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16071ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16072ED.seq | 3A | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16073ED.seq | 3A | 22/10/01 | 2bp_MOTU0051 | RF | RF | RF |
| 16074ED.seq | 3A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16075ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16076ED.seq | 3A | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 16078ED.seq | 3A | 22/10/01 | 2bp_MOTU0050 | RF | RF | RF |
| 16079ED.seq | 3A | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16080ED.seq | 3A | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16082ED.seq | 3A | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16083ED.seq | 3A | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16084ED.seq | 3A | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16101ED.seq | 3B | 22/10/01 | 2bp_MOTU0034 | EA | RF | RF |
| 16102ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16103ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16104ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16105ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16106ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16107ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16109ED.seq | 3B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16110ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16111ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16112ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16113ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16114ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16115ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16116ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16119ED.seq | 3B | 22/10/01 | 2bp_MOTU0010 | EA | RF | RF |
| 16120ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16121ED.seq | 3B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16124ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16125ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16126ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16127ED.seq | 3B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 16128ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16129ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16130ED.seq | 3B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16131ED.seq | 3B | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16132ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16133ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16135ED.seq | 3B | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 16136ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16137ED.seq | 3B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16138ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16139ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16140ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16141ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16142ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16143ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16144ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16145ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16147ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16148ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16149ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16150ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16151ED.seq | 3B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16152ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16153ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16154ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16155ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16156ED.seq | 3B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16157ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16158ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16160ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16161ED.seq | 3B | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16162ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16164ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16167ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16169ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16170ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16172ED.seq | 3B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16173ED.seq | 3B | 22/10/01 | 2bp_MOTU0104 | EA | RF | RF |
| 16174ED.seq | 3B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16177ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16178ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16179ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16180ED.seq | 3B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16182ED.seq | 3B | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16203ED.seq | 3C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16204ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16206ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16207ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16208ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16209ED.seq | 3C | 22/10/01 | 2bp_MOTU0112 | RF | RF | RF |
| 16210ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16211ED.seq | 3C | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16212ED.seq | 3C | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16213ED.seq | 3C | 22/10/01 | 2bp_MOTU0066 | RF | RF | RF |
| 16214ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16215ED.seq | 3C | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16217ED.seq | 3C | 22/10/01 | 2bp_MOTU0123 | RF | RF | RF |
| 16219ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16220ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16221ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16222ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16223ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16224ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16227ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16228ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16230ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16232ED.seq | 3C | 22/10/01 | 2bp_MOTU0009 | RF | RF | RF |
| 16234ED.seq | 3C | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16235ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16236ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16237ED.seq | 3C | 22/10/01 | 2bp_MOTU0066 | RF | RF | RF |
| 16238ED.seq | 3C | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16239ED.seq | 3C | 22/10/01 | 2bp_MOTU0066 | RF | RF | RF |
| 16240ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16241ED.seq | 3C | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16242ED.seq | 3C | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16243ED.seq | 3C | 22/10/01 | 2bp_MOTU0066 | RF | RF | RF |
| 16244ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16248ED.seq | 3C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16249ED.seq | 3C | 22/10/01 | 2bp_MOTU0060 | RF | RF | RF |
| 16251ED.seq | 3C | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16252ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16253ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16254ED.seq | 3C | 22/10/01 | 2bp_MOTU0106 | RF | RF | RF |
| 16255ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16257ED.seq | 3C | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16259ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16261ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16262ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16263ED.seq | 3C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16264ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16265ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16266ED.seq | 3C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16267ED.seq | 3C | 22/10/01 | 2bp_MOTU0118 | RF | RF | RF |
| 16270ED.seq | 3C | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 16271ED.seq | 3C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16302ED.seq | 3D | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16304ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16305ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16306ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16307ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16308ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16310ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16311ED.seq | 3D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16312ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16313ED.seq | 3D | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16314ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16315ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16316ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16318ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16319ED.seq | 3D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16320ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16321ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16322ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16323ED.seq | 3D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16324ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16325ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16326ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16327ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16328ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16329ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16331ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16332ED.seq | 3D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16334ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16337ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16338ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16339ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16340ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16341ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16342ED.seq | 3D | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16343ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16344ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16346ED.seq | 3D | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16347ED.seq | 3D | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16348ED.seq | 3D | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16349ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16350ED.seq | 3D | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 16351ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16352ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16353ED.seq | 3D | 22/10/01 | 2bp_MOTU0116 | RF | RF | RF |
| 16354ED.seq | 3D | 22/10/01 | 2bp_MOTU0021 | RF | RF | RF |
| 16355ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16356ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16357ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16358ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16359ED.seq | 3D | 22/10/01 | 2bp_MOTU0138 | RF | RF | RF |
| 16360ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16361ED.seq | 3D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16362ED.seq | 3D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16363ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16364ED.seq | 3D | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 16365ED.seq | 3D | 22/10/01 | 2bp_MOTU0057 | RF | RF | RF |
| 16366ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16367ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16368ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16370ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16371ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16372ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16373ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16377ED.seq | 3D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16378ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16379ED.seq | 3D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16401ED.seq | 3F | 22/10/01 | 2bp_MOTU0139 | RF | RF | RF |
| 16402ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16403ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16405ED.seq | 3F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16407ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16409ED.seq | 3F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16410ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16412ED.seq | 3F | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16413ED.seq | 3F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16414ED.seq | 3F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 16416ED.seq | 3F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16417ED.seq | 3F | 22/10/01 | 2bp_MOTU0065 | RF | RF | RF |
| 16418ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16420ED.seq | 3F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16422ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16423ED.seq | 3F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16424ED.seq | 3F | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 16425ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16427ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16428ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16430ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16431ED.seq | 3F | 22/10/01 | 2bp_MOTU0081 | RF | RF | RF |
| 16432ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16433ED.seq | 3F | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 16438ED.seq | 3F | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 16439ED.seq | 3F | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 16440ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16442ED.seq | 3F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 16443ED.seq | 3F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16444ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16446ED.seq | 3F | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16450ED.seq | 3F | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16451ED.seq | 3F | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 16452ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16454ED.seq | 3F | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16455ED.seq | 3F | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 16456ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16457ED.seq | 3F | 22/10/01 | 2bp_MOTU0001 | RF | RF | RF |
| 16458ED.seq | 3F | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16459ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16460ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16461ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16462ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16463ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16465ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16466ED.seq | 3F | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 16467ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16468ED.seq | 3F | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 16469ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16470ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16471ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16472ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16473ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16475ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16476ED.seq | 3F | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16477ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16478ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16479ED.seq | 3F | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16480ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16481ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16482ED.seq | 3F | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16483ED.seq | 3F | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 16484ED.seq | 3F | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16502ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16503ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16504ED.seq | 4B | 22/10/01 | 2bp_MOTU0041 | EA | RF | RF |
| 16506ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16507ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16508ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16509ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16511ED.seq | 4B | 22/10/01 | 2bp_MOTU0092 | EA | RF | RF |
| 16512ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16513ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16515ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16516ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16520ED.seq | 4B | 22/10/01 | 2bp_MOTU0010 | EA | RF | RF |
| 16522ED.seq | 4B | 22/10/01 | 2bp_MOTU0134 | EA | RF | RF |
| 16523ED.seq | 4B | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 16524ED.seq | 4B | 22/10/01 | 2bp_MOTU0038 | EA | RF | RF |
| 16525ED.seq | 4B | 22/10/01 | 2bp_MOTU0036 | EA | RF | RF |
| 16526ED.seq | 4B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16528ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16529ED.seq | 4B | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 16530ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16532ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16534ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16535ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16536ED.seq | 4B | 22/10/01 | 2bp_MOTU0077 | EA | RF | RF |
| 16537ED.seq | 4B | 22/10/01 | 2bp_MOTU0079 | EA | RF | RF |
| 16538ED.seq | 4B | 22/10/01 | 2bp_MOTU0003 | EA | RF | RF |
| 16540ED.seq | 4B | 22/10/01 | 2bp_MOTU0112 | EA | RF | RF |
| 16541ED.seq | 4B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16542ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16543ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16546ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16547ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16548ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16549ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16550ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16552ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16554ED.seq | 4B | 22/10/01 | 2bp_MOTU0095 | EA | RF | RF |
| 16555ED.seq | 4B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16556ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16557ED.seq | 4B | 22/10/01 | 2bp_MOTU0028 | EA | RF | RF |
| 16558ED.seq | 4B | 22/10/01 | 2bp_MOTU0038 | EA | RF | RF |
| 16559ED.seq | 4B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16561ED.seq | 4B | 22/10/01 | 2bp_MOTU0130 | EA | RF | RF |
| 16562ED.seq | 4B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16563ED.seq | 4B | 22/10/01 | 2bp_MOTU0121 | EA | RF | RF |
| 16564ED.seq | 4B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16566ED.seq | 4B | 22/10/01 | 2bp_MOTU0015 | EA | RF | RF |
| 16567ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16568ED.seq | 4B | 22/10/01 | 2bp_MOTU0010 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16573ED.seq | 4B | 22/10/01 | 2bp_MOTU0005 | EA | RF | RF |
| 16574ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16577ED.seq | 4B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16578ED.seq | 4B | 22/10/01 | 2bp_MOTU0055 | EA | RF | RF |
| 16580ED.seq | 4B | 22/10/01 | 2bp_MOTU0074 | EA | RF | RF |
| 16582ED.seq | 4B | 22/10/01 | 2bp_MOTU0046 | EA | RF | RF |
| 16583ED.seq | 4B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 16603ED.seq | 4C | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16604ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16605ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16607ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16608ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16609ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16610ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16611ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16612ED.seq | 4C | 22/10/01 | 2bp_MOTU0012 | EA | RF | RF |
| 16614ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16616ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16617ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16618ED.seq | 4C | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 16619ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16622ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16624ED.seq | 4C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 16625ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16626ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16627ED.seq | 4C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16628ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16629ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16630ED.seq | 4C | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 16631ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16633ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16635ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16636ED.seq | 4C | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16637ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16638ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16639ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16641ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16642ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16643ED.seq | 4C | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16645ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16646ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16647ED.seq | 4C | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16649ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16650ED.seq | 4C | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 16651ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16652ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16653ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16654ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16655ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16657ED.seq | 4C | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 16658ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16659ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16660ED.seq | 4C | 22/10/01 | 2bp_MOTU0073 | EA | RF | RF |
| 16662ED.seq | 4C | 22/10/01 | 2bp_MOTU0031 | EA | RF | RF |
| 16663ED.seq | 4C | 22/10/01 | 2bp_MOTU0075 | EA | RF | RF |
| 16665ED.seq | 4C | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 16666ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16667ED.seq | 4C | 22/10/01 | 2bp_MOTU0111 | EA | RF | RF |
| 16668ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16669ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16670ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16672ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16673ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16675ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16677ED.seq | 4C | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 16678ED.seq | 4C | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16679ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16680ED.seq | 4C | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16681ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 16682ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16683ED.seq | 4C | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16685ED.seq | 4C | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16686ED.seq | 4C | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16702ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16703ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16704ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16705ED.seq | 4D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16706ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16707ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16711ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16712ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16718ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16719ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16720ED.seq | 4D | 22/10/01 | 2bp_MOTU0020 | RF | RF | RF |
| 16721ED.seq | 4D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16722ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16723ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16724ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16725ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16726ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16728ED.seq | 4D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16729ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16732ED.seq | 4D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16733ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16734ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16736ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16737ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16738ED.seq | 4D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16739ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16740ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16741ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16742ED.seq | 4D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16743ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16744ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16745ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16746ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16747ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16748ED.seq | 4D | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 16749ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16750ED.seq | 4D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16751ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16753ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16754ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16755ED.seq | 4D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16756ED.seq | 4D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16757ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16761ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16762ED.seq | 4D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16801ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16802ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16803ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16805ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16806ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16807ED.seq | 4E | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 16808ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16809ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16810ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16811ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16812ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16813ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16814ED.seq | 4E | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 16815ED.seq | 4E | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 16816ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16817ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16818ED.seq | 4E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16819ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16820ED.seq | 4E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 16821ED.seq | 4E | 22/10/01 | 2bp_MOTU0132 | RF | RF | RF |
| 16822ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16823ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16824ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16825ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16826ED.seq | 4E | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 16827ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16828ED.seq | 4E | 22/10/01 | 2bp_MOTU0042 | RF | RF | RF |
| 16829ED.seq | 4E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16830ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16831ED.seq | 4E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 16832ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16834ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16835ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16836ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16837ED.seq | 4E | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 16838ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16839ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16841ED.seq | 4E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 16842ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16843ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16844ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16845ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16846ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16847ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16848ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16849ED.seq | 4E | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 16850ED.seq | 4E | 22/10/01 | 2bp_MOTU0083 | RF | RF | RF |
| 16851ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16852ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16855ED.seq | 4E | 22/10/01 | 2bp_MOTU0115 | RF | RF | RF |
| 16856ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16857ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16858ED.seq | 4E | 22/10/01 | 2bp_MOTU0136 | RF | RF | RF |
| 16859ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16860ED.seq | 4E | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 16861ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16862ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16863ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16864ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16865ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16866ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16867ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16868ED.seq | 4E | 22/10/01 | 2bp_MOTU0088 | RF | RF | RF |
| 16869ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16870ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16871ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16873ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16874ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16875ED.seq | 4E | 22/10/01 | 2bp_MOTU0031 | RF | RF | RF |
| 16876ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16877ED.seq | 4E | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 16878ED.seq | 4E | 22/10/01 | 2bp_MOTU0022 | RF | RF | RF |
| 16879ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16880ED.seq | 4E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 16881ED.seq | 4E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 16902ED.seq | 4F | 22/10/01 | 2bp_MOTU0059 | EA | RF | RF |
| 16903ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16905ED.seq | 4F | 22/10/01 | 2bp_MOTU0009 | EA | RF | RF |
| 16907ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16909ED.seq | 4F | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 16910ED.seq | 4F | 22/10/01 | 2bp_MOTU0017 | EA | RF | RF |
| 16911ED.seq | 4F | 22/10/01 | 2bp_MOTU0024 | EA | RF | RF |
| 16912ED.seq | 4F | 22/10/01 | 2bp_MOTU0017 | EA | RF | RF |
| 16914ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16915ED.seq | 4F | 22/10/01 | 2bp_MOTU0047 | EA | RF | RF |
| 16918ED.seq | 4F | 22/10/01 | 2bp_MOTU0117 | EA | RF | RF |
| 16919ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16920ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16921ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16922ED.seq | 4F | 22/10/01 | 2bp_MOTU0094 | EA | RF | RF |
| 16923ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16925ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16926ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16927ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16928ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16936ED.seq | 4F | 22/10/01 | 2bp_MOTU0066 | EA | RF | RF |
| 16937ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16939ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16941ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16942ED.seq | 4F | 22/10/01 | 2bp_MOTU0083 | EA | RF | RF |
| 16943ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16944ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16946ED.seq | 4F | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 16947ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16948ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16949ED.seq | 4F | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16951ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16952ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16953ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16955ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16956ED.seq | 4F | 22/10/01 | 2bp_MOTU0021 | EA | RF | RF |
| 16957ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16959ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16960ED.seq | 4F | 22/10/01 | 2bp_MOTU0017 | EA | RF | RF |
| 16961ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16962ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16963ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16964ED.seq | 4F | 22/10/01 | 2bp_MOTU0117 | EA | RF | RF |
| 16966ED.seq | 4F | 22/10/01 | 2bp_MOTU0045 | EA | RF | RF |
| 16967ED.seq | 4F | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16968ED.seq | 4F | 22/10/01 | 2bp_MOTU0026 | EA | RF | RF |
| 16970ED.seq | 4F | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 16971ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16973ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16974ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16975ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16977ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16978ED.seq | 4F | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 16979ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16980ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16981ED.seq | 4F | 22/10/01 | 2bp_MOTU0018 | EA | RF | RF |
| 16983ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 16984ED.seq | 4F | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17104ED.seq | 5B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 17105ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17106ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17107ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17108ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17109ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17110ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17112ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17113ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17114ED.seq | 5B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 17115ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17116ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17117ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17118ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17119ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17120ED.seq | 5B | 22/10/01 | 2bp_MOTU0004 | EA | RF | RF |
| 17121ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17122ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17123ED.seq | 5B | 22/10/01 | 2bp_MOTU0020 | EA | RF | RF |
| 17124ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17125ED.seq | 5B | 22/10/01 | 2bp_MOTU0014 | EA | RF | RF |
| 17126ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17127ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17128ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17129ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17130ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17131ED.seq | 5B | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 17132ED.seq | 5B | 22/10/01 | 2bp_MOTU0063 | EA | RF | RF |
| 17133ED.seq | 5B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 17134ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17135ED.seq | 5B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 17136ED.seq | 5B | 22/10/01 | 2bp_MOTU0063 | EA | RF | RF |
| 17137ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17138ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17140ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17141ED.seq | 5B | 22/10/01 | 2bp_MOTU0050 | EA | RF | RF |
| 17143ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17144ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17145ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17146ED.seq | 5B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 17147ED.seq | 5B | 22/10/01 | 2bp_MOTU0117 | EA | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17148ED.seq | 5B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 17149ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17150ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17151ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17153ED.seq | 5B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 17154ED.seq | 5B | 22/10/01 | 2bp_MOTU0034 | EA | RF | RF |
| 17156ED.seq | 5B | 22/10/01 | 2bp_MOTU0083 | EA | RF | RF |
| 17157ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17158ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17159ED.seq | 5B | 22/10/01 | 2bp_MOTU0034 | EA | RF | RF |
| 17160ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17161ED.seq | 5B | 22/10/01 | 2bp_MOTU0007 | EA | RF | RF |
| 17162ED.seq | 5B | 22/10/01 | 2bp_MOTU0011 | EA | RF | RF |
| 17163ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17164ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17165ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17167ED.seq | 5B | 22/10/01 | 2bp_MOTU0034 | EA | RF | RF |
| 17168ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17169ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17170ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17172ED.seq | 5B | 22/10/01 | 2bp_MOTU0006 | EA | RF | RF |
| 17173ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17174ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17175ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17176ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17177ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17178ED.seq | 5B | 22/10/01 | 2bp_MOTU0034 | EA | RF | RF |
| 17179ED.seq | 5B | 22/10/01 | 2bp_MOTU0014 | EA | RF | RF |
| 17180ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17181ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17182ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17184ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17185ED.seq | 5B | 22/10/01 | 2bp_MOTU0002 | EA | RF | RF |
| 17186ED.seq | 5B | 22/10/01 | 2bp_MOTU0016 | EA | RF | RF |
| 17188ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17189ED.seq | 5B | 22/10/01 | 2bp_MOTU0133 | EA | RF | RF |
| 17190ED.seq | 5B | 22/10/01 | 2bp_MOTU0063 | EA | RF | RF |
| 17191ED.seq | 5B | 22/10/01 | 2bp_MOTU0008 | EA | RF | RF |
| 17205ED.seq | 5C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17208ED.seq | 5C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17224ED.seq | 5C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17227ED.seq | 5C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17228ED.seq | 5C | 22/10/01 | 2bp_MOTU0012 | RF | RF | RF |
| 17231ED.seq | 5C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17234ED.seq | 5C | 22/10/01 | 2bp_MOTU0012 | RF | RF | RF |
| 17235ED.seq | 5C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 17236ED.seq | 5C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17238ED.seq | 5C | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 17245ED.seq | 5C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17246ED.seq | 5C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 17247ED.seq | 5C | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 17249ED.seq | 5C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17251ED.seq | 5C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 17252ED.seq | 5C | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17253ED.seq | 5C | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17255ED.seq | 5C | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 17257ED.seq | 5C | 22/10/01 | 2bp_MOTU0016 | RF | RF | RF |
| 17259ED.seq | 5C | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 17262ED.seq | 5C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17264ED.seq | 5C | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17265ED.seq | 5C | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 17266ED.seq | 5C | 22/10/01 | 2bp_MOTU0027 | RF | RF | RF |
| 17301ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17302ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17303ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17304ED.seq | 5D | 22/10/01 | 2bp_MOTU0056 | RF | RF | RF |
| 17305ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17307ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17309ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17311ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17312ED.seq | 5D | 22/10/01 | 2bp_MOTU0022 | RF | RF | RF |
| 17314ED.seq | 5D | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 17315ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17316ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17317ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17318ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17319ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17320ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17322ED.seq | 5D | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 17324ED.seq | 5D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17325ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17326ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17327ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17328ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17329ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17330ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17334ED.seq | 5D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17338ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17339ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17341ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17342ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17343ED.seq | 5D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 17344ED.seq | 5D | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 17346ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17347ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17348ED.seq | 5D | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17349ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17350ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17351ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17354ED.seq | 5D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17355ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17356ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17357ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17358ED.seq | 5D | 22/10/01 | 2bp_MOTU0025 | RF | RF | RF |
| 17360ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17361ED.seq | 5D | 22/10/01 | 2bp_MOTU0071 | RF | RF | RF |
| 17362ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17363ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17365ED.seq | 5D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17367ED.seq | 5D | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17368ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17369ED.seq | 5D | 22/10/01 | 2bp_MOTU0037 | RF | RF | RF |
| 17370ED.seq | 5D | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17371ED.seq | 5D | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17403ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17404ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17405ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17406ED.seq | 5E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17407ED.seq | 5E | 22/10/01 | 2bp_MOTU0032 | RF | RF | RF |
| 17408ED.seq | 5E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 17409ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17410ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17411ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17412ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17413ED.seq | 5E | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 17415ED.seq | 5E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17416ED.seq | 5E | 22/10/01 | 2bp_MOTU0026 | RF | RF | RF |
| 17417ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17418ED.seq | 5E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17419ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17420ED.seq | 5E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 17421ED.seq | 5E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17422ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17423ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17424ED.seq | 5E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 17425ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17426ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17427ED.seq | 5E | 22/10/01 | 2bp_MOTU0011 | RF | RF | RF |
| 17428ED.seq | 5E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 17429ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17430ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17433ED.seq | 5E | 22/10/01 | 2bp_MOTU0033 | RF | RF | RF |
| 17434ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17435ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17436ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17437ED.seq | 5E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17438ED.seq | 5E | 22/10/01 | 2bp_MOTU0007 | RF | RF | RF |
| 17439ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17440ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17441ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17443ED.seq | 5E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 17444ED.seq | 5E | 22/10/01 | 2bp_MOTU0010 | RF | RF | RF |
| 17448ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17457ED.seq | 5E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17458ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17459ED.seq | 5E | 22/10/01 | 2bp_MOTU0008 | RF | RF | RF |
| 17460ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17461ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17462ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17464ED.seq | 5E | 22/10/01 | 2bp_MOTU0014 | RF | RF | RF |
| 17465ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17466ED.seq | 5E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17467ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17468ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17469ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17470ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17471ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17472ED.seq | 5E | 22/10/01 | 2bp_MOTU0124 | RF | RF | RF |
| 17473ED.seq | 5E | 22/10/01 | 2bp_MOTU0049 | RF | RF | RF |
| 17474ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17476ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17477ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |
| 17478ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17479ED.seq | 5E | 22/10/01 | 2bp_MOTU0018 | RF | RF | RF |
| 17480ED.seq | 5E | 22/10/01 | 2bp_MOTU0002 | RF | RF | RF |
| 17484ED.seq | 5E | 22/10/01 | 2bp_MOTU0004 | RF | RF | RF |
| 17486ED.seq | 5E | 22/10/01 | 2bp_MOTU0006 | RF | RF | RF |

Table 2: named sequences from other datasets.

| **Organism name** | **Acc. no. / source** |
|---|---|
| Acrobeles_ciliatus | AF202148 |
| Acrobeloides_bodenheimeri | AF202162 |
| Alaimus_sp | P. De Ley, pers. comm. |
| Allodorylaimus_sp | P. De Ley, pers. comm. |
| Anaplectus_sp | J. Vanfleteren, pers. comm. |
| Aphelenchoides_fragariae | J. Vanfleteren, pers. comm. |
| Aphelenchus_avenae | AF036586 |
| Aporcelaimellus_obtusicaudatus2 | M. Blaxter, pers. comm. |
| Boleodorus_sp_clone2 | J. Vanfleteren, pers. comm. |
| Bunonema_franzi | P. De Ley, pers. comm. |
| Bursaphelenchus_sp | AF037369 |
| Caenorhabditis_briggsae | U13929 |
| Caenorhabditis_drosophilae | AF083025 |
| Caenorhabditis_elegans | X03680 |
| Caenorhabditis_sonorae | AF083026 |
| Caenorhabditis_vulgaris | U13931 |
| Cephalobus_cubaensis | AF202161 |
| Cephalobus_oryzae | AF034390 |
| Chiloplectus_sp | J. Vanfleteren, pers. comm. |
| Clarkus_sp | J. Vanfleteren, pers. comm. |
| Daptonema_procerus | AF047889 |
| Desmodora_ovigera | Y16913 |

| | |
|---|---|
| Ditylenchus_angustus | P. De Ley, pers. comm. |
| Dorylaimida_B | J. Vanfleteren, pers. comm. |
| Ecphyadophoridae_sp | J. Vanfleteren, pers. comm. |
| Eubostrichus_topiarius | Y16917 |
| Eudorylaimus_carteri | P. De Ley, pers. comm. |
| Geocenamus_quadrifer | J. Vanfleteren, pers. comm. |
| Globodera_pallida | AF036592 |
| Gordius_aquaticus | X80233 |
| Helicotylenchus_dihystera | P. De Ley, pers. comm. |
| Longidorus_elongatus | AF036594 |
| Meloidogyne_incognita | U81578 |
| Mesodorylaimus_bastiani | P. De Ley, pers. comm. |
| Mononchus_truncatus | J. Vanfleteren, pers. comm. |
| Mylonchulus_arenicolus | AF036596 |
| Panagrellus_redivivus | AF036599 |
| Panagrolaimus_sp_PS1159 | U81579 |
| Paractinolaimus_macrolaimus | J. Vanfleteren, pers. comm. |
| Paratylenchus_dianthus | J. Vanfleteren, pers. comm. |
| Pellioditis_marina | AF083021 |
| Pellioditis_mediterranea | AF083020 |
| Pellioditis_typica | U13933 |
| Plectus_acuminatus | AF037628 |
| Plectus_aquatilis | AF036602 |
| Plectus_minimus | P. De Ley, pers. comm. |
| Praeacanthonchus_sp | AF036612 |
| Pratylenchoides_magnicauda | AF202157 |
| Pratylenchoides_ritteri | J. Vanfleteren, pers. comm. |
| Pratylenchus_goodeyi | P. De Ley, pers. comm. |
| Pratylenchus_thornei | J. Vanfleteren, pers. comm. |
| Prionchulus_muscorum | P. De Ley, pers. comm. |
| Prismatolaimus_intermedius | AF036603 |
| Pristionchus_lheritieri | AF036640 |
| Pseudacrobeles_variabilis | AF202150 |
| Pungentus_sp | J. Vanfleteren, pers. comm. |
| Rotylenchus_robustus | P. De Ley, pers. comm. |
| Steinernema_affine | M. Blaxter, pers. comm. |
| Steinernema_carpocapsae | AF036604 |
| Steinernema_feltiae | M. Blaxter, pers. comm. |
| Subanguina_radicicola | AF202164 |
| Teratocephalus_lirellius | AF036607 |
| Trichodorus_primitivus | AF036609 |
| Tridentulus_sp | P. De Ley, pers. comm. |
| Tripyla_sp | M. Blaxter, pers. comm. |
| Tylencholaimus_sp | J. Vanfleteren, pers. comm. |
| Wilsonema_schuurmansstekhoveni | P. De Ley, pers. comm. |
| Xiphinema_rivesi | AF036610 |
| Zygotylenchus_guevarae | J. Vanfleteren, pers. comm. |

# Appendix 2: List of MOTU

This section lists the set of 140 2bp_MOTU, as shown in figure 4.7.1. Each MOTU has been putatively assigned to a taxonomic group on the basis of similarity to a known sequence; some sequences have been placed to genus, some to family, and others only to order/suborder, depending on the degree of similarity to the comparison sequence. Also included is the abundance of each MOTU in the full set of 2039 sequences tested.

| MOTU name | Genus | Family | Order/Suborder | Abundance |
|---|---|---|---|---|
| 2bp_MOTU0001 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0002 | Helicotylenchus | Hoplolaimidae | Tylenchina | 835 |
| 2bp_MOTU0003 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0004 | Tripyla | Tripylidae | Enoplida | 117 |
| 2bp_MOTU0005 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0006 | Aporcelaimellus | Aporcelaimidae | Dorylaimida | 318 |
| 2bp_MOTU0007 | unknown | Mononchidae | Mononchida | 26 |
| 2bp_MOTU0008 | Allodorylaimus | Qudsianematidae | Dorylaimida | 67 |
| 2bp_MOTU0009 | unknown | Plectidae | Plectida | 14 |
| 2bp_MOTU0010 | unknown | Qudsianematidae | Dorylaimida | 29 |
| 2bp_MOTU0011 | Eudorylaimus | Qudsianematidae | Dorylaimida | 56 |
| 2bp_MOTU0012 | unknown | unknown | Enoplida | 16 |
| 2bp_MOTU0013 | Aphelenchoides | Aphelenchoididiae | Tylenchina | 1 |
| 2bp_MOTU0014 | Mesodorylaimus | Thornenematidae | Dorylaimida | 4 |
| 2bp_MOTU0015 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0016 | Pungentus | Noriidae | Dorylaimida | 79 |
| 2bp_MOTU0017 | Paratylenchus | Paratylenchidae | Tylenchina | 5 |
| 2bp_MOTU0018 | unknown | unknown | Dorylaimida | 65 |
| 2bp_MOTU0019 | Caenorhabditis | Rhabditidae | Rhabditina | 4 |
| 2bp_MOTU0020 | unknown | Mononchidae | Mononchida | 30 |
| 2bp_MOTU0021 | unknown | Cephalobidae | Tylenchina | 3 |
| 2bp_MOTU0022 | unknown | Qudsianematidae | Dorylaimida | 3 |
| 2bp_MOTU0023 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0024 | Teratocephalus | Teratocephalidae | Chromadorida | 1 |
| 2bp_MOTU0025 | Pratylenchoides | Pratylenchidae | Tylenchina | 12 |
| 2bp_MOTU0026 | unknown | Mononchidae | Mononchida | 53 |
| 2bp_MOTU0027 | Rotylenchus | Hoplolaimidae | Tylenchina | 42 |
| 2bp_MOTU0028 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0029 | unknown | unknown | Dorylaimida | 4 |
| 2bp_MOTU0030 | Alaimus | Alaimidae | Enoplida | 7 |
| 2bp_MOTU0031 | Alaimus | Alaimidae | Enoplida | 19 |
| 2bp_MOTU0032 | unknown | unknown | Dorylaimida | 7 |
| 2bp_MOTU0033 | Steinernema | Steinernematidae | Tylenchina | 15 |
| 2bp_MOTU0034 | unknown | Plectidae | Plectida | 9 |
| 2bp_MOTU0035 | Panagrolaimus | Panagrolaimidae | Rhabditina | 1 |
| 2bp_MOTU0036 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0037 | Acrobeloides | Cephalobidae | Tylenchina | 18 |
| 2bp_MOTU0038 | Helicotylenchus | Hoplolaimidae | Tylenchina | 2 |
| 2bp_MOTU0039 | Boleodorus | Tylenchidae | Tylenchina | 1 |
| 2bp_MOTU0040 | Teratocephalus | Teratocephalidae | Chromadorida | 2 |
| 2bp_MOTU0041 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0042 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |

| | | | | |
|---|---|---|---|---|
| 2bp_MOTU0043 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0044 | Prismatolaimus | Prismatolaimidae | Enoplida | 1 |
| 2bp_MOTU0045 | Subanguina | Anguinidae | Tylenchina | 2 |
| 2bp_MOTU0046 | Helicotylenchus | Hoplolaimidae | Tylenchina | 8 |
| 2bp_MOTU0047 | unknown | Cephalobidae | Tylenchina | 1 |
| 2bp_MOTU0048 | Pellioditis | Rhabditidae | Rhabditina | 8 |
| 2bp_MOTU0049 | unknown | unknown | Dorylaimida | 14 |
| 2bp_MOTU0050 | Subanguina | Anguinidae | Tylenchina | 2 |
| 2bp_MOTU0051 | unknown | unknown | Dorylaimida | 6 |
| 2bp_MOTU0052 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0053 | Subanguina | Anguinidae | Tylenchina | 1 |
| 2bp_MOTU0054 | Prismatolaimus | Prismatolaimidae | Enoplida | 1 |
| 2bp_MOTU0055 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0056 | unknown | Mononchidae | Mononchida | 1 |
| 2bp_MOTU0057 | unknown | unknown | Dorylaimida | 2 |
| 2bp_MOTU0058 | Steinernema | Steinernematidae | Tylenchina | 3 |
| 2bp_MOTU0059 | Paratylenchus | Paratylenchidae | Tylenchina | 1 |
| 2bp_MOTU0060 | Paratylenchus | Paratylenchidae | Tylenchina | 6 |
| 2bp_MOTU0061 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0062 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0063 | unknown | Mononchidae | Mononchida | 3 |
| 2bp_MOTU0064 | Boleodorus | Tylenchidae | Tylenchina | 1 |
| 2bp_MOTU0065 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0066 | Helicotylenchus | Hoplolaimidae | Tylenchina | 5 |
| 2bp_MOTU0067 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0068 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0069 | unknown | Mononchidae | Mononchida | 1 |
| 2bp_MOTU0070 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0071 | Pratylenchoides | Pratylenchidae | Tylenchina | 6 |
| 2bp_MOTU0072 | Subanguina | Anguinidae | Tylenchina | 1 |
| 2bp_MOTU0073 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0074 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0075 | Rotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0076 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0077 | Alaimus | Alaimidae | Enoplida | 1 |
| 2bp_MOTU0078 | unknown | Mononchidae | Mononchida | 2 |
| 2bp_MOTU0079 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0080 | Teratocephalus | Teratocephalidae | Chromadorida | 1 |
| 2bp_MOTU0081 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0082 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0083 | unknown | Cephalobidae | Tylenchina | 3 |
| 2bp_MOTU0084 | unknown | Plectidae | Plectida | 1 |
| 2bp_MOTU0085 | Paratylenchus | Paratylenchidae | Tylenchina | 1 |
| 2bp_MOTU0086 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0087 | Subanguina | Anguinidae | Tylenchina | 1 |
| 2bp_MOTU0088 | unknown | Plectidae | Plectida | 1 |
| 2bp_MOTU0089 | Pratylenchus | Pratylenchidae | Tylenchina | 1 |
| 2bp_MOTU0090 | Subanguina | Anguinidae | Tylenchina | 1 |
| 2bp_MOTU0091 | Globodera | Heteroderidae | Tylenchina | 2 |
| 2bp_MOTU0092 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0093 | Tripyla | Tripylidae | Enoplida | 2 |
| 2bp_MOTU0094 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0095 | Bunonema | Bunonematidae | Rhabditina | 1 |
| 2bp_MOTU0096 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0097 | unknown | unknown | Dorylaimida | 2 |
| 2bp_MOTU0098 | Paratylenchus | Paratylenchidae | Tylenchina | 2 |
| 2bp_MOTU0099 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0100 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0101 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0102 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |

| | | | | |
|---|---|---|---|---|
| 2bp_MOTU0103 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0104 | unknown | Mononchidae | Mononchida | 1 |
| 2bp_MOTU0105 | unknown | Plectidae | Plectida | 1 |
| 2bp_MOTU0106 | unknown | unknown | Monhysterida | 1 |
| 2bp_MOTU0107 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0108 | Tripyla | Tripylidae | Enoplida | 1 |
| 2bp_MOTU0109 | Subanguina | Anguinidae | Tylenchina | 2 |
| 2bp_MOTU0110 | Subanguina | Anguinidae | Tylenchina | 1 |
| 2bp_MOTU0111 | Panagrolaimus | Panagrolaimidae | Rhabditina | 1 |
| 2bp_MOTU0112 | Tridentulus | Monhysteridae | Monhysterida | 2 |
| 2bp_MOTU0113 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0114 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0115 | unknown | Plectidae | Plectida | 1 |
| 2bp_MOTU0116 | Boleodorus | Tylenchidae | Tylenchina | 1 |
| 2bp_MOTU0117 | unknown | Mononchidae | Mononchida | 4 |
| 2bp_MOTU0118 | unknown | Cephalobidae | Tylenchina | 1 |
| 2bp_MOTU0119 | Pratylenchus | Pratylenchidae | Tylenchina | 1 |
| 2bp_MOTU0120 | Rotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0121 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0122 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0123 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0124 | unknown | unknown | Monhysterida | 2 |
| 2bp_MOTU0125 | Alaimus | Alaimidae | Enoplida | 2 |
| 2bp_MOTU0126 | unknown | Plectidae | Plectida | 1 |
| 2bp_MOTU0127 | unknown | unknown | Tylenchina | 1 |
| 2bp_MOTU0128 | Pellioditis | Rhabditidae | Rhabditina | 1 |
| 2bp_MOTU0129 | Caenorhabditis | Rhabditidae | Rhabditina | 1 |
| 2bp_MOTU0130 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0131 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0132 | Pellioditis | Rhabditidae | Rhabditina | 1 |
| 2bp_MOTU0133 | unknown | Cephalobidae | Tylenchina | 1 |
| 2bp_MOTU0134 | Pratylenchoides | Pratylenchidae | Tylenchina | 1 |
| 2bp_MOTU0135 | unknown | unknown | Dorylaimida | 1 |
| 2bp_MOTU0136 | Prismatolaimus | Prismatolaimidae | Enoplida | 1 |
| 2bp_MOTU0137 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |
| 2bp_MOTU0138 | Tridentulus | Monhysteridae | Monhysterida | 1 |
| 2bp_MOTU0139 | unknown | Alaimidae | Enoplida | 1 |
| 2bp_MOTU0140 | Helicotylenchus | Hoplolaimidae | Tylenchina | 1 |

# Appendix 3: Perl Scripts

This section includes the Perl source code for a number of scripts which were written as part of this project. All text beginning with a hash symbol ("#") are comments, and not part of the program itself.

## 1. define_motu.pl

This script takes a set of sequences and assigns them to MOTU, based on a user-defined level of sequence identity. It takes as its input a set of sequence files, in FASTA format, which must end in the extension ".seq" and must be placed in a directory called "sequences". Each time the script is run, the processing order of these sequences is randomised.

Each sequence in turn is compared using BLAST to a set of sequences which have already been assigned MOTU names. The first sequence to go through is therefore searched against an empty database and gets no matches; it is automatically assigned to "MOTU0001". The second sequence is then searched against a database containing only the first sequence; if it matches (i.e. the number of base differences between them is equal to or fewer than the number entered by the user at the start), the second sequence is also assigned to MOTU0001; if it does not match, it is assigned to MOTU0002. The third sequence is then searched against both of the first two, and so on. This process continues until every sequence has been assigned to a MOTU.

Each distinct MOTU is named by assigning it a unique number. Additionally, the program includes the number of base differences used as a part of the MOTU name: for example, if 3 bases is chosen, each MOTU is named "3bp_MOTU...". It is also possible for the user to add a prefix to each MOTU name by entering the desired text at the command prompt when the script is run. This feature is useful if, for example, several runs are being carried out on the same set of sequences. By entering:

```
> define_MOTU.pl [space] run001_ [enter]
```

each MOTU in that run will be given a name beginning "run001_2bp_MOTU..." (assuming 2 bp is chosen as the threshold). Therefore, the set of MOTU defined by each subsequent run can be given a unique set of names (e.g. run002..., run003..., etc.) allowing the output of each run to be easily distinguished.

```
#!/usr/bin/perl
#
#Searches a set of sequences in a random order and places into MOTU

use File::stat;

# The user is first asked to specify the number of base differences
# allowed between sequences within a MOTU
```

```
print "How many bases to define MOTU?\t";

chomp($BASES = <STDIN>);

$PREFIX = $ARGV[0];
$MOTUNAME = $PREFIX.$BASES."bp_MOTU";
$DATABASE = $MOTUNAME;
$DATABASE_LOCATION = "./$DATABASE";
$| = 1;      # make STDOUT print buffer flush immediately

$rootdir = "./OUT";
$seq_files = "./sequences";          # Directory to find the sequence files
$seq_files1 = "./sequences_done";    # Directory to place used sequence files

# Set up the directories and relevant files if not already present

$INDNUMBER=0;
$TOPINDNUMBER=0;

if(!stat(OUT))   { system("mkdir OUT"); }
if(!stat($seq_files1))   { system("mkdir $seq_files1"); }
if(!stat($DATABASE)) {system("touch $DATABASE"); }

else
#If an existing MOTU file is present, continue the numbering system from this file
{
    open(fh,"<$DATABASE");
    while($line=<fh>)
    {
        if($line=~/$MOTUNAME(\d+)/ && $1 > $INDNUMBER)
        {
            $INDNUMBER=$1; $INDNUMBER=~s/^0+//;
            print "New Indnum is $INDNUMBER\n";
        }
    }
    close(fh);
}

# Create the master list for cluster generation

system("rm master") if -e "master";
system("rm randomnumbers") if -e "randomnumbers";

#First, count the number of .seq files in the current directory

opendir(SEQDIR, "$seq_files");
$filecount=0;
while (defined($file= readdir(SEQDIR)))
{
    if($file=~/seq/) { ++$filecount; }
}
print "$filecount files found \n";
closedir(SEQDIR);

# Next bit creates an array of random integers between 1 and
# the number of files found by the previous section

$numbercount = 0;
until ($numbercount == $filecount)

{
    $randomnumber=0;
    while ($randomnumber == 0 || $matchflag == 1)
    {
        $randomnumber = int(rand $filecount + 1);
        $matchflag = 0;
        foreach(0..$numbercount)
        {
            if ($numberlist[$_] == $randomnumber)
            #Makes sure the same number is not generated twice
            {
                $matchflag = 1;
```

```
                last;
            }
        }
    }

$numberlist[$numbercount] = $randomnumber ;
++$numbercount;

}

open(RANDOMFILE,">>randomnumbers");

#Print the list of random numbers to a file in case needed for future reference

foreach(@numberlist) { print RANDOMFILE "$_,"; }

print "@numberlist\n";

#Create a new array with each filename in a position corresponding to the random numbers.

opendir(SEQDIR, "$seq_files");
$a=0;

while (defined($file= readdir(SEQDIR)))

{
    if($file=~/seq/)
    {
        $filelist[$numberlist[$a]-1]=$file;
        ++$a;
    }
}

closedir(SEQDIR);
open(MASTER,">>master");
foreach(@filelist) { print MASTER "$_,"; }
print "@filelist\n";
print "$a files processed\n";


@searchlist = @filelist;

# Open a log file for the session.  Use the date to give each logfile a distinctive name.

($min,$hr,$mday,$yr) = (localtime)[1,2,3,5];
$min = "0" . $min if $min < 10;
$hr = "0" . $hr if $hr < 10;
$mon =(Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec)[(localtime)[4]];
$log = "$rootdir/logfile_$hr:${min}_$mday$mon$yr";
open(LOG, "| tee -a $log");
select(LOG); $| = 1;

# Look at the sequence files, one file at a time.
# You have to be in the correct source directory to do this.

$j = 0;
SEARCH:

while($file = shift(@searchlist))
{

    ++$j;
    print LOG "$j. Blasting $file\n";

#Blast the file against the set of sequences already assigned to MOTU.

    system "/usr/ncbi/bin/blastall -p blastn -d $DATABASE -i $seq_files/$file -G 2 -E 1 -v 100
-F F -o $rootdir/$file.out >> $log";
    open(OUTFILE,"<$rootdir/$file.out") || die "can't open $rootdir/$file.out\n";

#The following variables are set to zero each time a new file is searched

    $match_flag=0 ;
```

```
    $identities=0;
    $match_length=0;
    $gaps=0;
    $query='';
    $subject='';
    $qlength=0;
    $ns=0;
    $corrected_match_length=0;

    while (<OUTFILE>)   #search the BLAST output
    {

        if(/>.+(\d\d\d\d)$/)   #find the index number of the top hit MOTU
        {
            $INDNUMBER=$1; $INDNUMBER=~s/^0+//;
            print "** $INDNUMBER\n";
        }

        if(/Identities\s\=\s(\d+)\/(\d+)/)   #get number of identities and match length

        {
            $identities = $1 ; $match_length = $2 ;
            if(/Gaps\s=\s(\d+)/) {$gaps = $1;}
        }

# The next 2 lines read the 'subject' and 'query' sequences into
# the string variables $subject and $query...

        if(/Query:\s\d+\s+(.+)\s\d+/) { $query .= $1 ; }

        if(/Sbjct:\s\d+\s+(.+)\s\d+/) { $subject .= $1 ; }

# When the length of sequence read equals the match length already defined, it
# has finished reading the sequence and leaves the loop. Also need to make sure
# that $query and $subject are the same length.

        if(length($query)>0 && length($query)==$match_length &&
length($query)==length($subject))
        { last; }

    }

    close(OUTFILE);

    $ns = &count_ns($query,$subject);   #subroutine gets the number of Ns

    $corrected_match_length = $match_length - $gaps - $ns ;

    if($corrected_match_length != 0) #to avoid division by zero...
    {
        $percentID = ($identities/$corrected_match_length)*100;
        $base_diff = $corrected_match_length-$identities;
    }

# Display the output on screen

    print "Identities=$identities\n";
    print "Match length=$match_length\n";
    print "Gaps=$gaps\nNs=$ns\n";
    print "Corrected match length=$corrected_match_length\n";
    print "Percent ID=$percentID\nNo. base differences=$base_diff\n";

# The next section tests whether the number of base differences is less
# than or equal to the number of bases defined by the user at the beginning
# (if so, the current sequence is assigned the same MOTU name as its match);
# or greater, in which case the current sequence must be given a new MOTU name.

    if($base_diff <= $BASES && $match_length >= 300) {$match_flag=1;}

    else {$match_flag=0;}

    if($match_flag==1)    #Got a match - use existing MOTU name
    { $rightstring = $MOTUNAME.&index($INDNUMBER); }
```

```
    else
# No match - make new MOTU name by adding 1 to the highest number used so far

    {
        ++$INDNUMBER;
        if ($INDNUMBER <= $TOPINDNUMBER)

        {
            $rightstring = $MOTUNAME.&index(++$TOPINDNUMBER);
        }

        else
        {
            $TOPINDNUMBER = $INDNUMBER;
            $rightstring = $MOTUNAME.&index($INDNUMBER);
        }
    }


    &newfile($file, $rightstring);
    print LOG "$file assigned to $rightstring\n\n";
    system "cat $seq_files1/$file >> $DATABASE_LOCATION";
    system "/usr/ncbi/bin/formatdb -i $DATABASE_LOCATION -p F";

}

print "$j sequences processed; $TOPINDNUMBER $MOTUNAME defined.\n";


################################  subroutines  ################################


sub index
# Put zeros in front of the index to make it four digits long.
#
{
    local($i) = @_;
    if     ($i < 10)          { $index = "000$i"; }
    elsif ($i < 100)          { $index = "00$i"; }
    elsif ($i < 1000)         { $index = "0$i"; }
    elsif ($i < 10000)        { $index = "$i"; }
    else                      { die "index out of range\n"; }
}

sub newfile
# Write a new file containing the sequence name, the MOTU to which it
# has been assigned, and the sequence itself
{
    ($file, $newstring) = @_;
    rename("$seq_files/$file", "$seq_files1/$file.old");
    open(OLDFILE,"$seq_files1/$file.old");
    open(FILE,">$seq_files1/$file");
    while(<OLDFILE>) {
        print FILE unless /^>/;
        if (/^>/) {
            chop($original = $_);
            print FILE "$original $newstring\n";
        }
    }
    close(OLDFILE);
    close(FILE);
}


sub count_ns
# This subroutine counts the number of Ns in a pair of sequences, so that
# this number can be subtracted from the match length. It reads both the
# query and subject sequences. Where it finds an N, it must also check that
# the matching position in the other sequence is not also an N or a gap,
# so that the same position is not subtracted twice.

{
```

```
    $ncount=0; $ns=0; $q=0; $s=0;
    @query_array = split('',$query);
    @subject_array = split('',$subject);

    foreach $qbase (@query_array)
    {
        if ($qbase eq ('n' || 'N'))
        {
            unless ($subject_array[$q] =~ /-|n|N/) {++$ncount;}
            print "$qbase, $subject_array[$q]\n";
        }
        ++$q;
    }

    foreach $sbase (@subject_array)
    {
        if ($sbase eq ('n' || 'N'))
        {
            if ($query_array[$s] =~ /-|n|N/) {++$ncount;}
            print "$sbase, $query_array[$s]\n";
        }
        ++$s;
    }

    $ns=$ncount;
}
```

## 2. primer_trim.pl

This script was used to trim each sequence to a constant length by searching for a specified pair of conserved "tag" sequences found at the 5' and 3' ends in most nematode SSUs, and deleting everything outside those sites (the name of this script is a slight misnomer: it was originally written to search for primer sites, but any pair of sequences can be searched for - they do not have to be primer sites, and indeed the sequences in this version of the script are not). Also, any sequence for which a match is not found is flagged, so that nonmatching sequences can be picked out and examined individually, and if necessary, manually trimmed to an appropriate length.

```
#!/usr/bin/perl
#
# primer_trim.pl
#
# Finds primer sites within sequence files and trims sequence outside
# the primer sites.

system ("rm matches") if -e "matches";
system ("rm non_matches") if -e "non_matches";
system ("rm trimmed_table") if -e "trimmed_table";

$fwd_primer='TGCATG' ;  # conserved 3' tag
$rev_primer='CAAT\wTAAA' ;  # conserved 5' tag

opendir(DIR, "./");

while (defined($file=readdir(DIR)))
{

# Reset the relevant variables at the start of each cycle

    $seq_name=$file;
    $sequence='';
    $preprimer='';
    $postprimer='';

    if($file=~/\.seq/)
    {

        open(SEQFILE,"<$file") || die "can't open $file\n";

        while (<SEQFILE>)
        {

            #if(/(^>.+\.seq)/) { $seq_name=$1 ; }

            if(/(^\w+)/) { $sequence.=$1 ; }
            #reads sequence into variable $sequence

        }

        close(SEQFILE);

        if($sequence =~ /(\w*$fwd_primer)/)
        {

            $preprimer = $1;
            $sequence =~ s/$preprimer// ;
            #find sequence before fwd primer, then delete it
            $matches .= "3' tag match found for $seq_name\n";
```

```
        }

        else
        {
            $non_matches .= "No 3' tag match for $seq_name\n";
            $seq_name.="_no3match";
        }

        if($sequence =~ /$rev_primer(\w*)/)
        {

            $postprimer = $1;
            $sequence =~ s/$postprimer// ;
            #find sequence after rev primer, then delete it
            $matches .= "5' tag match found for $seq_name\n";
        }

        else
        {
            $non_matches .= "No 5' tag match for $seq_name\n";
            $seq_name.="_no5match";
        }

        $seq_length=length($sequence);

        system("touch trimmed_table\n");

        open(OUTFILE,">>trimmed_table");

        print OUTFILE ">$seq_name $seq_length\n$sequence\n";
    }

}

close(OUTFILE);

system("touch non_matches\n");

system("touch matches\n");

open(NMFILE,">>non_matches");

print NMFILE "$non_matches" ;

close(NMFILE);

open(MFILE,">>matches");

print MFILE "$matches" ;

close(MFILE);

closedir(DIR);
```

## 3. column_var.pl

This script was used to examine the positions of variable sites in an alignment of sequences. An alignment was first generated using ClustalX software. As part of its output this program is able to save a column scores file, in which all of the bases in each column (aligned site) appear on a single line separated by spaces. This file is read by the **column_var.pl** script, which then determines for each aligned site (1) the consensus (i.e. most common) base, and (2) the variability of that site, i.e. the number of bases in that column which are *not* the consensus base (and also are not gaps or Ns). If every base in every sequence at a given site is identical, that site has a variability of zero. The output of this script could then be plotted as a histogram as shown in Figure 5.2.1.

```
#!/usr/bin/perl
#
#column_var.pl

$i=0;

open(INFILE, $ARGV[0]);

#INFILE should be a clustalw/x columns scores file

while (<INFILE>)
{
    if (/^(\D+)\s+\d+/) { push @column, $1; }
}

# Each element of the array @column is a string containing all the
# letters in that column, separated by spaces.

foreach(0..$#column)
{
    $count{'A'}=0;
    $count{'G'}=0;
    $count{'C'}=0;
    $count{'T'}=0;
    $count{'N'}=0;
    $count{'-'}=0;
    $column_number=$_;

#Turn each column from a string to an array by splitting on spaces

    @column_array=split('\s',$column[$_]);
    foreach $letter (@column_array)
    {   #count each letter
        if($letter =~ /A|G|C|T|N|-/) { ++$count{$letter}; }
        else { die "error: illegal character in column $column_number\n"; }
    }

    print "$_:\t@column_array\n";
    print "A: $count{'A'}\t";
    print "G: $count{'G'}\t";
    print "C: $count{'C'}\t";
    print "T: $count{'T'}\t";
    print "N: $count{'N'}\t";
    print "gaps: $count{'-'}\n";

    @sorted = sort by_number keys %count;
    $consensus[$_]=$sorted[0];
#Finds the consensus, i.e. most common letter

    if ($consensus[$_] eq 'A') { $var[$_] = $count{G}+$count{C}+$count{T};}
```

```
    if ($consensus[$_] eq 'G') { $var[$_] = $count{A}+$count{C}+$count{T};}
    if ($consensus[$_] eq 'C') { $var[$_] = $count{G}+$count{A}+$count{T};}
    if ($consensus[$_] eq 'T') { $var[$_] = $count{G}+$count{C}+$count{A};}

    print "Consensus: $consensus[$_]\tvariability: $var[$_]\n";
}

system ("rm column_scores") if -e "column_scores";
system("touch column_var\n");
open(OUTFILE,">>column_var");

# Creates a tab-delimited table of column number, consensus
# base, and variability

foreach(0..$column_number)
{
    $j=$_+1;
    print OUTFILE "$j\t$consensus[$_]\t$var[$_]\n";
}

sub by_number { $count{$b} <=> $count{$a} }
```

## 4. columns_exclude_var.pl

This script was used to examine how variable sites are linked within an alignment. Like **column_var.pl**, this script determines, for each site, the consensus base and the variability; however, for all sites whose variability is greater than zero, it also determines which particular sequences within that column differ from the consensus base. It then temporarily deletes those sequences (i.e. rows) from the entire alignment, carries out the analysis again, and counts how much overall variability has been removed by eliminating those sequences. This procedure is repeated for every variable column in the alignment.

```
#!/usr/bin/perl
#
#columns_exclude_var.pl

$i=0;

open(INFILE, $ARGV[0]);

while (<INFILE>)
{
    if (/^(\D+)\s+\d+/)
    {
        $column[$i]=$1;
        ++$i;
    }
}

foreach(0..$#column)
{
    $column_number=$_;
    undef %count;
    @column_array=split('\s',$column[$_]);

    foreach $letter (@column_array)
    {
        if($letter =~ /A|G|C|T|N|-/) { ++$count{$letter}; }
        else { die "error: illegal character in column $column_number\n"; }
    }

    @sorted = sort by_number keys %count;
    $consensus[$_]=$sorted[0];

    if ($consensus[$_] eq 'A') { $var[$_] = $count{G}+$count{C}+$count{T}; }
    if ($consensus[$_] eq 'G') { $var[$_] = $count{A}+$count{C}+$count{T}; }
    if ($consensus[$_] eq 'C') { $var[$_] = $count{G}+$count{A}+$count{T}; }
    if ($consensus[$_] eq 'T') { $var[$_] = $count{G}+$count{C}+$count{A}; }

    foreach(0..$#column_array)
    {
        unless ($column_array[$_] =~ /$consensus[$column_number]|N|-/)
        { $nonmatches[$column_number] .= "$_," ; }
    }
}

system ("rm comparisons") if -e "comparisons";
system("touch comparisons\n");
open(OUTFILE,">>comparisons");

system ("rm summary") if -e "summary";
system("touch summary\n");
open(OUTFILE2,">>summary");

foreach(0..$#column)
{
```

```
    if($var[$_] > 0)
    {
        $total_old=0;
        $total_new=0;
        $total_diff=0;
        $j=$_+1;
        @var_array=split(',',$nonmatches[$_]);
        print OUTFILE "$j:\tSequence(s) @var_array removed.\n";

        foreach(0..$#column)
        {
            $k=$_+1;
            undef %count;
            @reduced_array=split('\s',$column[$_]);

            foreach(@var_array) { $reduced_array[$_]='-'; }

            foreach $letter (@reduced_array) { ++$count{$letter}; }

            @sorted = sort by_number keys %count;
            $new_consensus[$_]=$sorted[0];

            if($new_consensus[$_] ne $consensus[$_])
            { die "something odd has happened"; }

            if ($consensus[$_] eq 'A') { $new_var[$_] = $count{G}+$count{C}+$count{T};}
            if ($consensus[$_] eq 'G') { $new_var[$_] = $count{A}+$count{C}+$count{T};}
            if ($consensus[$_] eq 'C') { $new_var[$_] = $count{G}+$count{A}+$count{T};}
            if ($consensus[$_] eq 'T') { $new_var[$_] = $count{G}+$count{C}+$count{A};}

            $difference[$_]=$var[$_]-$new_var[$_];

            $total_old=$total_old+$var[$_];
            $total_new=$total_new+$new_var[$_];
            $total_diff=$total_diff+$difference[$_];

            print OUTFILE"$k:$var[$_]\t$new_var[$_]\t$difference[$_]\n";
        }
        print OUTFILE "\n";
        print OUTFILE2 "$j\t$var[$_]\t$total_diff\t$nonmatches[$_]\n";
        print "$j\t$var[$_]\t$total_diff\t$nonmatches[$_]\n";
    }
}

sub by_number { $count{$b} <=> $count{$a} }
```

## 5. div_table.pl

This script takes as its input the set of MOTU assignments produced by **define_MOTU.pl**, and outputs a table showing, for each sampling unit (plot and sampling date) which MOTU are present and the number of members in each MOTU. This script is able to incorporate multiple runs of **define_MOTU.pl**, and produce a single table combining information from all runs.

```
#!/usr/bin/perl
#

%site = (

          101 => 'mixed',
          102 => '4D',
          104 => '1F_June',
          105 => '1F_June',
          106 => '1F_June',
          107 => '1F_June',
          108 => '1F_June',
          109 => '1F_June',
          110 => '1F_June',
          111 => '1F_June',
          112 => '2B_June',
          113 => '2B_June',
          114 => '2B_June',
          115 => '2B_June',
          116 => '2B_June',
          117 => '2B_June',
          118 => '2B_June',
          119 => '2B_June',
          120 => '3D_June',
          121 => '3D_June',
          122 => '3D_June',
          123 => '3D_June',
          124 => '3D_June',
          125 => '3D_June',
          126 => '3D_June',
          127 => '3D_June',
          128 => '4D_June',
          129 => '4D_June',
          130 => '4D_June',
          131 => '4D_June',
          132 => '4D_June',
          133 => '4D_June',
          134 => '4D_June',
          135 => '4D_June',
          136 => '5A_June',
          137 => '5A_June',
          138 => '5A_June',
          139 => '5A_June',
          140 => '5A_June',
          141 => '5A_June',
          142 => '5A_June',
          143 => '5A_June',
          150 => '1A_Oct',
          151 => '1B_Oct',
          152 => '1C_Oct',
          153 => '1E_Oct',
          154 => '1F_Oct',
          155 => '2A_Oct',
          156 => '2B_Oct',
          157 => '2C_Oct',
          158 => '2D_Oct',
```

```
          159 => '2E_Oct',
          160 => '3A_Oct',
          161 => '3B_Oct',
          162 => '3C_Oct',
          163 => '3D_Oct',
          164 => '3F_Oct',
          165 => '4B_Oct',
          166 => '4C_Oct',
          167 => '4D_Oct',
          168 => '4E_Oct',
          169 => '4F_Oct',
          170 => '5A_Oct',
          171 => '5B_Oct',
          172 => '5C_Oct',
          173 => '5D_Oct',
          174 => '5E_Oct',
          );

system ("rm tempfile") if -e "tempfile";

opendir(DIR,"./");

while (defined($file= readdir(DIR)))
{
    undef %taxon_set;

    open(MOTUFILE,$file);

    while(<MOTUFILE>)
    {
        if(/>(\d+ED).+\s(.+)\n/)
        {
            $seq_name=$1;
            $taxon_name=$2;
            if($seq_name=~/(\d\d\d).+/) { $source=$1 };
            $taxon_set{$site{$source}}.=$taxon_name.",";
        }
    }
    close(MOTUFILE);

    open(ABUNDFILE,"$file.abund");
    while(<ABUNDFILE>)
    {
        if(/^(.+)\t(\d+)$/)
        { $abundance{$1}=$2; }
    }
    close(ABUNDFILE);

    open(TEMPFILE,">>tempfile");

    undef %taxon_string;
    undef %count_string;
    undef %abundance_string;

    foreach $sample (sort keys %taxon_set)
    {
        undef %count;
        $taxon_richness{$sample}=0;
        $uniques{$sample}=0;
        $unique_nonsingles{$sample}=0;

        @taxon_array=split(',',$taxon_set{$sample});
        $total_indiv{$sample}=scalar(@taxon_array);

        foreach $taxon(@taxon_array) {  ++$count{$taxon}; }

        foreach(sort keys %count)
        {
            ++$taxon_richness{$sample};
            $taxon_string{$sample}.=$_.",";
            $count_string{$sample}.=$count{$_}.",";
            $abundance_string{$sample}.=$abundance{$_}.",";
            if($count{$_}==$abundance{$_})
```

```
            {
                ++$uniques{$sample};
                if($count{$_} > 1) { ++$unique_nonsingles{$sample}; }
            }
        }
    }

    foreach $sample (sort keys %taxon_set)
    {
        print TEMPFILE "Sample $sample\n";
        print TEMPFILE "Taxa $taxon_string{$sample}\n";
        print TEMPFILE "Abundances $count_string{$sample}\n";
        print TEMPFILE "Overall $abundance_string{$sample}\n";
        print TEMPFILE "Uniques $uniques{$sample}\t$unique_nonsingles{$sample}\n";
    }
}

close(TEMPFILE);

undef %uniques;
undef %unique_nonsingles;

open(TEMPFILE,"tempfile");

while(<TEMPFILE>)
{
    if(/Sample\s(.+)\n/) { $sample=$1; }

    if(/Taxa\s(.+)\n/) { $taxon_superset{$sample}.=$1.";"; }

    if(/Abundances\s(.+)\n/) { $number_superset{$sample}.=$1.";"; }

    if(/Uniques\s(\d+)\t(\d+)\n/)
    { $uniques{$sample}.=$1.";"; $unique_nonsingles{$sample}.=$2.";"; }
}

close(TEMPFILE);

system ("rm tempfile");

system ("rm div_table") if -e "div_table";

open(OUTFILE,">>div_table");

foreach (sort keys %taxon_superset)
{
    @taxon_superarray=split(';',$taxon_superset{$_});
    @number_superarray=split(';',$number_superset{$_});
    @uniques_array=split(';',$uniques{$_});
    @unique_nonsingles_array=split(';',$unique_nonsingles{$_});

    print OUTFILE "Sample $_\n";

    foreach(0..$#taxon_superarray)
    {
        print OUTFILE "$taxon_superarray[$_]\n";
        print OUTFILE "$number_superarray[$_]\n";
        print OUTFILE "$uniques_array[$_]\t$unique_nonsingles_array[$_]\n";
    }
}
```

# 6. div_indices.pl

This script takes the output of **div_table.pl** and calculates a set of diversity indices for each sampling event, producing separate output for each MOTU run. The parameters determined are: number of individuals in the sample, absolute number of taxa, number of unique taxa, number of unique non-singleton taxa, Shannon index, Simpson index and Berger-Parker dominance index.

```
#!/usr/bin/perl
#

open(FILE,"$ARGV[0]");

while(<FILE>)
{
    if(/Sample\s(.+)\n/) { $source=$1; }

    if(/^(run\d+)_/) { $run=$1; }

    $name=$source."_".$run;

    if(/^(\d+,.*)\n/) { $abundance{$name}=$1; }
    if(/^(\d+)\t(\d+)/) { $uniques{$name}=$1; $unique_nonsingles{$name}=$2; }
}

close(FILE);

open(OUTFILE,">>div_indices");

foreach(sort keys %abundance)
{
    @array=split(",",$abundance{$_});
    $total_ind=0;
    $total_tax=scalar(@array);
    $shannon=0;
    $simpson=0;
    undef $neglogSI;

    foreach(@array) { $total_ind=$total_ind+$_; }

    foreach $n (@array)
    {
        $pi=$n/$total_ind;
        $pi_logpi=-($pi*log($pi));
        $shannon=$shannon+$pi_logpi;

        $nsquared=$n*$n;
        $Nsquared=$total_ind*$total_ind;
        unless($Nsquared == 1)
        {
            $ratio=($nsquared-$n)/($Nsquared-$total_ind);
            $simpson=$simpson+$ratio;
        }
    }

    unless($simpson == 0) { $neglogSI=-log($simpson); }

    @sorted = sort {$b<=>$a} @array;
    $dominant=$sorted[0];
    $berger=$dominant/$total_ind;

    print OUTFILE "$_\t$total_ind\t$total_tax\t";
    print OUTFILE "$uniques{$_}\t$unique_nonsingles{$_}\t";
    print OUTFILE "$shannon\t$neglogSI\t$berger\n";
    #print "$_\t$total_ind\t$total_tax\t$shannon\t$neglogSI\t$berger\n";
}
```

# 7. div_summary.pl

This script is designed to summarise the output of **div_indices.pl**. When 100 runs of **define_MOTU.pl** were carried out, the output of **div_indices.pl** was a very large file, containing diversity parameters for all sites and sampling dates 100 times over. This was summarised into a table containing, for each site and date, a mean, minimum and maximum value for each of the diversity parameters among the 100 runs. This is the information shown in Tables 6.2.1, 6.3.1 and 6.3.2.

```
#!/usr/bin/perl
#

open(INFILE,$ARGV[0]);

while(<INFILE>)
{
    if(/^(.+_.+)_.+\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(.+)\t(.+)\t(.+)\n/)
    {
        $sample_name=$1;
        $nind{$sample_name}=$2;
        $ntax{$sample_name}.=$3.",";
        $uniques{$sample_name}.=$4.",";
        $unique_nonsingles{$sample_name}.=$5.",";
        $shannon{$sample_name}.=$6.",";
        $simpson{$sample_name}.=$7.",";
        $dominance{$sample_name}.=$8.",";
    }
}

close(INFILE);

system ("rm div_summary") if -e "div_summary";
open(OUTFILE,">>div_summary");

foreach(sort keys %shannon)
{
    $ntax_total=0;
    $uniques_total=0;
    $unique_nonsingles_total=0;
    $shannon_total=0;
    $simpson_total=0;
    $dominance_total=0;

    @ntax_array=split(',',$ntax{$_});
    @uniques_array=split(',',$uniques{$_});
    @unique_nonsingles_array=split(',',$unique_nonsingles{$_});
    @shannon_array=split(',',$shannon{$_});
    @simpson_array=split(',',$simpson{$_});
    @dominance_array=split(',',$dominance{$_});

    foreach(@ntax_array)
    { $ntax_total=$ntax_total+$_; }

    foreach(@uniques_array)
    { $uniques_total=$uniques_total+$_; }

    foreach(@unique_nonsingles_array)
    { $unique_nonsingles_total=$unique_nonsingles_total+$_; }

    foreach(@shannon_array)
    { $shannon_total=$shannon_total+$_; }

    foreach(@simpson_array)
    { $simpson_total=$simpson_total+$_; }
```

```
    foreach(@dominance_array)
    { $dominance_total=$dominance_total+$_; }

    @ntax_sorted=sort {$a <=> $b} @ntax_array;
    @uniques_sorted=sort {$a <=> $b} @uniques_array;
    @unique_nonsingles_sorted=sort {$a <=> $b} @unique_nonsingles_array;
    @shannon_sorted=sort {$a <=> $b} @shannon_array;
    @simpson_sorted=sort {$a <=> $b} @simpson_array;
    @dominance_sorted=sort {$a <=> $b} @dominance_array;

    $ntax_mean=$ntax_total/scalar(@ntax_array);
    $ntax_min=$ntax_sorted[0];
    $ntax_max=$ntax_sorted[$#ntax_sorted];

    $uniques_mean=$uniques_total/scalar(@uniques_array);
    $uniques_min=$uniques_sorted[0];
    $uniques_max=$uniques_sorted[$#uniques_sorted];

    $unique_nonsingles_mean=$unique_nonsingles_total/scalar(@unique_nonsingles_array);
    $unique_nonsingles_min=$unique_nonsingles_sorted[0];
    $unique_nonsingles_max=$unique_nonsingles_sorted[$#unique_nonsingles_sorted];

    $shannon_mean=$shannon_total/scalar(@shannon_array);
    $shannon_min=$shannon_sorted[0];
    $shannon_max=$shannon_sorted[$#shannon_sorted];

    $simpson_mean=$simpson_total/scalar(@simpson_array);
    $simpson_min=$simpson_sorted[0];
    $simpson_max=$simpson_sorted[$#simpson_sorted];

    $dominance_mean=$dominance_total/scalar(@dominance_array);
    $dominance_min=$dominance_sorted[0];
    $dominance_max=$dominance_sorted[$#dominance_sorted];

    print OUTFILE "$_\t$nind{$_}\t$ntax_min\t$ntax_max\t$ntax_mean\t";
    print OUTFILE "$uniques_min\t$uniques_max\t$uniques_mean\t";
    print OUTFILE "$unique_nonsingles_min\t$unique_nonsingles_max\t$unique_nonsingles_mean\t";
    print OUTFILE "$shannon_min\t$shannon_max\t$shannon_mean\t";
    print OUTFILE "$simpson_min\t$simpson_max\t$simpson_mean\t";
    print OUTFILE "$dominance_min\t$dominance_max\t$dominance_mean\n";
}
```

# Appendix 4: Published Article

There follows a reprint of the paper:

Floyd, R., Abebe, E., Papert, A. and Blaxter, M. (2002). "Molecular barcodes for soil nematode identification." Molecular Ecology 11: 839-850.

This presents the data from the preliminary Sourhope survey, and information on the development of some of the methods. It is reprinted here with the permission of Blackwell Science Ltd, and of the co-authors.

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 *ExoI* 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

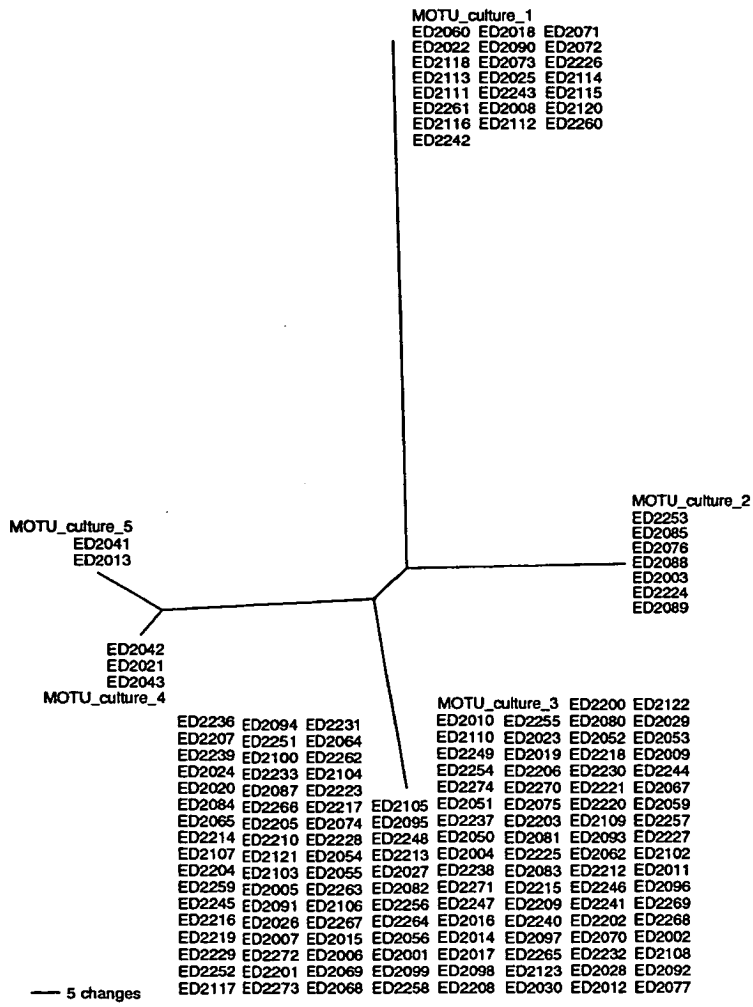



**Fig. 1** Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from cultured nematode isolates. One hundred and sixty-six sequences from nematode cultures initiated from single specimens were aligned and analysed as described in materials and methods. The analysis included 349 of the aligned 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.

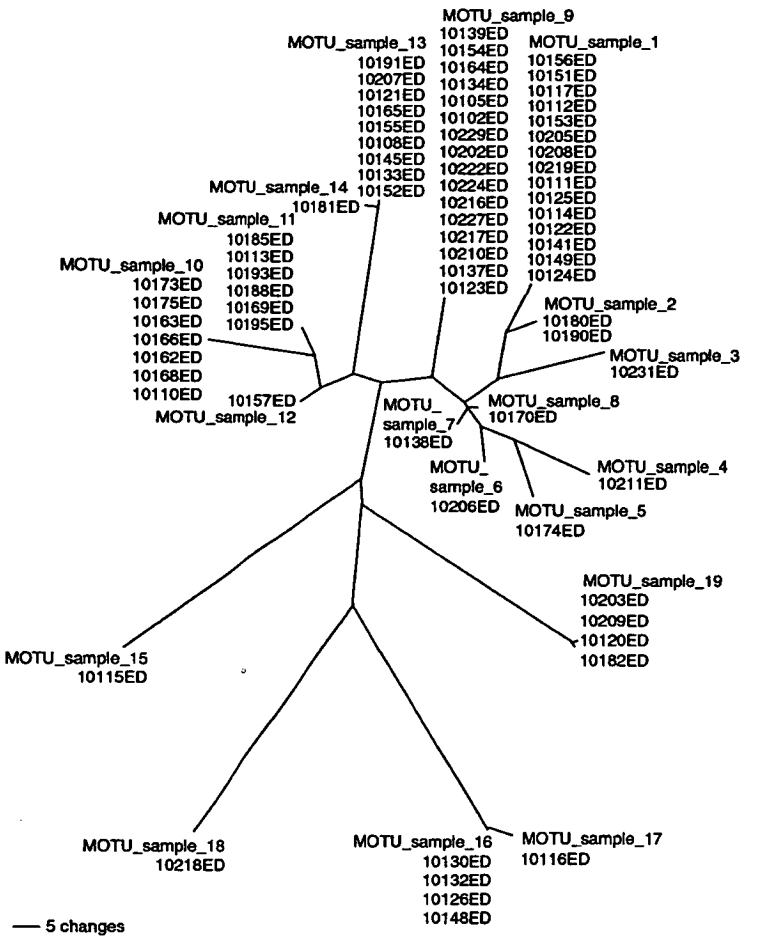



Fig. 2 Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from a random sample. Seventy-four sequences derived from single nematode specimens across the Sourhope field site were aligned and analysed as described in materials and methods. The analysis included 350 of the aligned nucleotides. 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

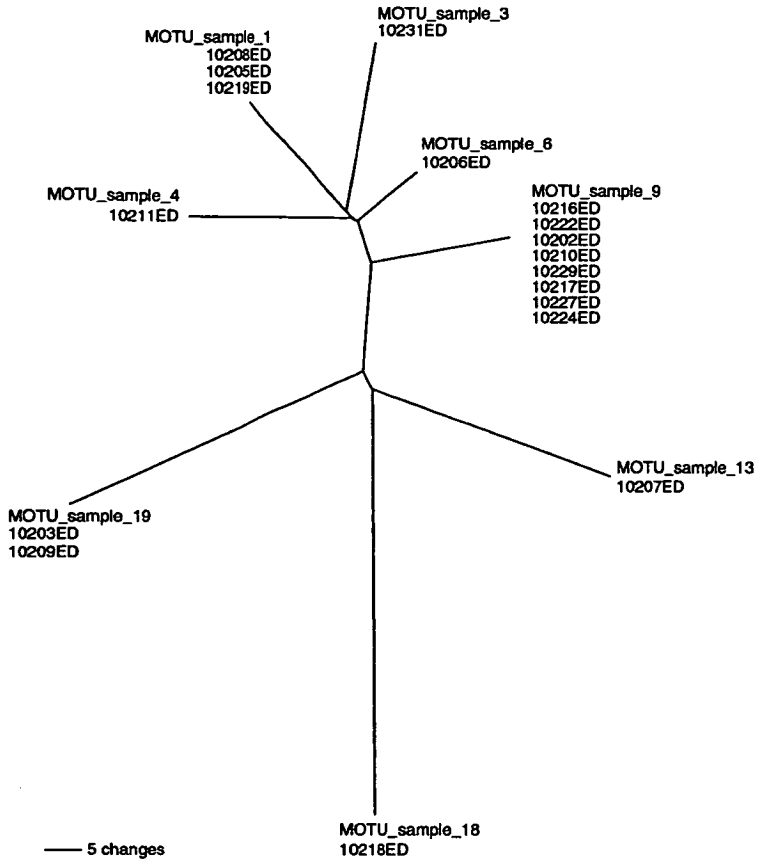



**Fig. 3** Unrooted phylogram of 5′ end small subunit ribosomal RNA sequences from a random sample from a single subplot. Eighteen sequences from a single subplot of the Sourhope field site (designated 4 DU) were aligned and analysed as described in materials and methods. The analysis included 396 of 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

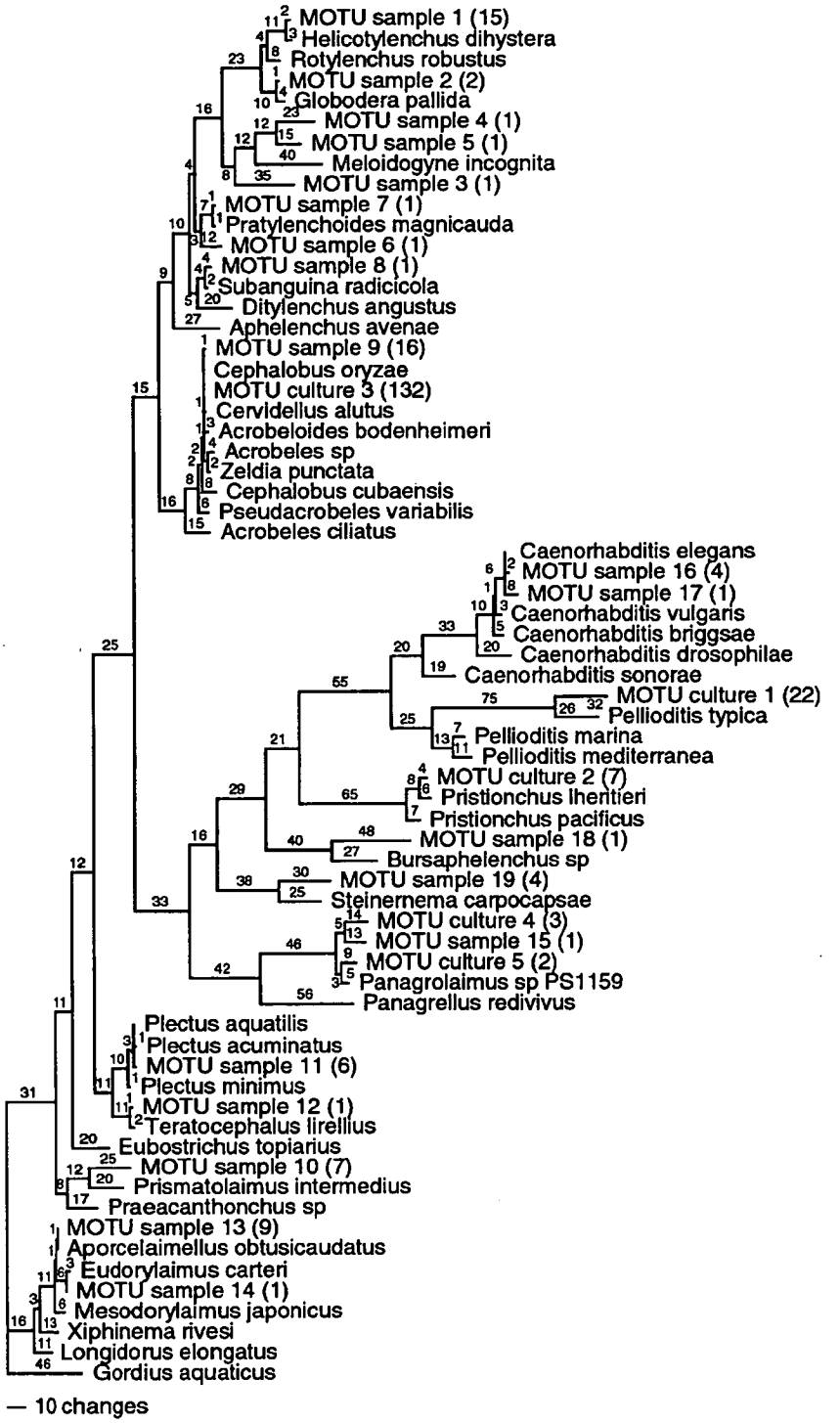



**Fig. 4** Phylogram, rooted using the nematomorph *Gordius aquaticus* as an outgroup, of a neighbour joining analysis of all 24 survey sequences from this study and a selected set of 43 sequences from identified taxa. The alignment included 554 characters. The alignment was subjected to NJ analysis 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.

---

# References


Adams, B. J. (2001). "The species delimitation uncertainty principle." Journal of Nematology **33**: 153-160.

Adams, B. J., Burnell, A. M. and Powers, T. O. (1998). "A phylogenetic analysis of the genus *Heterorhabditis* (Nemata: Rhabditidae) based on internal transcribed spacer 1 DNA sequence data." Journal of Nematology **30**: 22-29.

Aleshin, V. V., Kedrova, O. S., Milyutina, I. A., Vladychenskaya, N. S. and Petrov, N. B. (1998a). "Relationships among nematodes based on the analysis of 18S rRNA gene sequences: molecular evidence for a monophyly of chromadorian and secernentean nematodes." Russian Journal of Nematology **6**: 175-184.

Aleshin, V. V., Milyutina, I. A., Kedrova, O. S., Vladychenskaya, N. S. and Petrov, N. B. (1998b). "Phylogeny of Nematoda and Cephalorhyncha derived from 18S rDNA." Journal of Molecular Evolution **47**: 597-605.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**: 403-10.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**: 3389-3402.

Amaral Zettler, L. A., Gomez, F., Zettler, E., Keenan, B. G., Amils, R. and Sogin, M. L. (2002). "Eukaryotic diversity in Spain's River of Fire." Nature **417**: 137.

Anderson, I. C., Campbell, C. D. and Prosser, J. I. (2003). "Potential bias of fungal18S rDNA and internal transcribed spacer polyerase chain reaction primers for estimating fungal biodiversity in soil." Environmental Microbiology **5**: 36-47.

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O. and Zimmer, E. (1980). "Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and ape." Proceedings of the National Academy of Sciences USA **77**: 7323-7327.

Baird, S. M. and Bernard, E. C. (1984). "Nematode population and community dynamics in soybean-wheat cropping and tillage regimes." Journal of Nematology **16**: 379-386.

Barns, S. M., Takala, S. L. and Kuske, C. R. (1999). "Wide distribution and diversity of members of the bacterial kingdom *Acidobacterium* in the environment." Applied and Environmental Microbiology **65**: 1731-1737.

Beckenbach, K., Blaxter, M. and Webster, J. M. (1999). "Phylogeny of *Bursaphelenchus* species derived from analysis of ribosomal internal transcribed spacer DNA sequences." Nematology **1**: 539-548.

Berger, W. H. and Parker, F. L. (1970). "Diversity of planktonic foraminifera in deep-sea sediments." Science **168**: 1345-1347.

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

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

Blaxter, M. L., De Ley, P., Garey, J., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T. and Thomas, W. K. (1998). "A molecular evolutionary framework for the phylum Nematoda." Nature **392**: 71-75.

Bloemers, G. F., Hodda, M., Lambshead, P. J. D., Lawton, J. H. and Wanless, F. R. (1997). "The effects of forest disturbance on diversity of tropical soil nematodes." Oecologia **111**: 575-582.

Boag, B. and Yeates, G. W. (1998). "Soil nematode biodiversity in terrestrial ecosystems." Biodiversity and Conservation **7**: 617-630.

Bongers, T. (1990). "The maturity index: an ecological measure of environmental disturbance based on nematode species composition." Oecologia **83**: 14-19.

Bonnet, R., Suau, A., Doré, J., Gibson, G. R. and Collins, M. D. (2002). "Differences in rDNA libraries of faecal bacteria derived from 10- and 25- cycle PCRs." International Journal of Systematic and Evolutionary Microbiology **52**: 757-763.

Brooks, D. R. and McLennan, D. A. (1999). "Species: turning a conundrum into a research program." Journal of Nematology **31**: 117-133.

Bush, G. L. (1994). "Sympatric speciation in animals: new wine in old bottles." Trends in Ecology & Evolution **9**: 285-288.

Chao, A. (1984). "Nonparametric estimation of the number of classes in a population." Scandinavian Journal of Statistics **11**: 265-270.

Claridge, M. F., Dawah, H. A. and Wilson, M. R., Eds. (1997). Species: The Units of Biodiversity. Chapman & Hall

Cracraft, J. (1997). Species concepts in systematics and conservation biology - an ornithological viewpoint. In: Species: The Units of Biodiversity (M. F. Claridge, H. A. Dawah and M. R. Wilson, Eds.), pp. 325-339. Chapman & Hall.

Darwin, C. (1859). The Origin of Species. John Murray, London.

De Ley, P. (2000). "Lost in worm space: phylogeny and morphology as road maps to nematode diversity." Nematology **2**: 9-16.

De Ley, P. and Blaxter, M. L. (2001). Systematic position and phylogeny. In: The Biology of Nematodes (D. Lee, Ed.), pp. 1-30. Harwood Academic Publishers, Reading.

De Ley, P., Felix, M.-A., Frisse, L. M., Nadler, S. A., Sternberg, P. W. and Thomas, W. K. (1999). "Molecular and morphological characterisation of two reproductively isolated species with mirror-image anatomy (Nematoda: Cephalobidae)." Nematology **1(6)**: 591-612.

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

Dover, G. (1995). "A species definition: a functional approach." Trends in Ecology & Evolution **10**: 489-490.

Eggert, L. S., Rasner, C. A. and Woodruff, D. S. (2002). "The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers." Proc. R. Soc. Lond. B. **269**: 1993-2006.

Eldredge, N. and Cracraft, J. (1980). Phylogenetic Patterns and the Evolutionary Process. Columbia University Press

Ettema, C. H. (1998). "Soil nematode diversity: species coexistence and ecosystem function." Journal of Nematology **30**: 159-169.

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

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

Eyualem, A. and 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.

Eyualem, A., Mees, J. and Coomans, A. (2001). "Nematode communities of Lake Tana and other inland water bodies of Ethiopia." Hydrobiologica **462**: 41-73.

Felix, M.-A., De Ley, P., Sommer, R., Frisse, L., Nadler, S. J., Thomas, W. K., Vanfleteren, J. and Sternberg, P. W. (2000). "Evolution of vulva development in the Cephalobina (Nematoda)." Developmental Biology **221**: 68-86.

Fitch, D. H. A., Bugaj-Gaweda, B. and Emmons, S. W. (1995). "18S ribosomal RNA gene phylogeny for some Rhabditidae related to *Caenorhabditis*." Molecular Biology & Evolution **12**: 346-348.

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

Foucher, A. and Wilson, M. (2002). "Development of a polymerase chain reaction-based denaturing gradient gel electrophoresis technique to study nematode species biodiversity using the 18s rDNA gene." Molecular Ecology Notes **2**: 45-48.

Freckman, D. W., Ed. (1982). Nematodes in Soil Ecosystems. University of Texas Press, Austin.

Freckman, D. W. and Virginia, R. A. (1997). "Low-diversity Antarctic soil nematode communities: distribution and response to disturbance." Ecology **78**: 363-369.

Furlong, M. A., Singleton, D. R., Coleman, D. C. and Whitman, W. B. (2002). "Molecular and culture-based analyses of prokaryotic communities from an agricultural soil and the burrows and casts of the earthworm *Lumbricus rubellus*." Applied and Environmental Microbiology **68**: 1265-1279.

Futuyma, D. J. (1998). Evolutionary Biology, 3rd ed. Sinauer Associates

Godfray, H. C. J. (2002). "Challenges for taxonomy." Nature **417**: 17-19.

Groombridge, B. (1992). Global biodiversity. Chapman & Hall, London.

Großkopf, R., Stubner, S. and Liesack, W. (1998). "Novel Euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms." Applied and Environmental Microbiology **64**: 4983-4989.

Hagstrom, A., Pommier, T., Rohwe, F., Simu, K., Stolte, W., Svensson, D. and Zweifel, U. L. (2002). "Use of ribosomal DNA for delineation of marine bacterioplankton species." Applied and Environmental Microbiology **68**: 3628-3633.

Hägstrom, A., Pommier, T., Rohwe, F., Simu, K., Stolte, W., Svensson, D. and Zweifel, U. L. (2002). "Use of ribosomal DNA for delineation of marine bacterioplankton species." Applied and Environmental Microbiology **68**: 3628-3633.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and deWaard, J. R. (2003a). "Biological identifications through DNA barcodes." Proc. R. Soc. Lond. B. **270**: 313-321.

Hebert, P. D. N., Ratnasingham, S. and deWaard, J. R. (2003b). "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species." Proc. R. Soc. Lond. B. (Suppl.): DOI 10.1098/rsbl.2003.0025.

Hey, J. (2001). "The mind of the species problem." Trends in Ecology & Evolution **16**: 326-329.

Hillis, D. M. and Dixon, M. T. (1991). "Ribosomal DNA: molecular evolution and phylogenetic inference." Quarterly Review of Biology **66**: 411-453.

Hull, D. L. (1997). The ideal species concept - and why we can't get it. In: Species: The Units of Biodiversity (M. F. Claridge, H. A. Dawah and M. R. Wilson, Eds.), pp. 357-380. Chapman & Hall.

Johnson, S. R., Ferris, V. R. and Ferris, J. M. (1972). "Nematode community structure of forest woodlots. I. Relationships based on similarity coefficients of nematode species." Journal of Nematology **4**: 175-183.

Kampfer, S., Sturmbauer, C. and Ott, J. (1998). "Phylogenetic analysis of rDNA sequences from adenophorean nematodes and implications for the Adenophorea-Secernentea controversy." Invertebrate Biology **117**: 29-36.

Knowlton, N. and Weight, L. A. (1997). Species of marine invertebrates: a comparison of the biological and phylogenetic species concepts. In: Species: The Units of Biodiversity (M. F. Claridge, H. A. Dawah and M. R. Wilson, Eds.), pp. Chapman & Hall.

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

Lambshead, P. D. J. (2001). Marine nematode diversity. In: Nematology, Advances and Perspectives (Z. X. Chen, S. Y. Chen and D. W. Dickson, Eds.), pp. ACSE-TUP Book Series, San Francisco, USA.

Lambshead, P. J. D., Tietjen, J., Ferrero, T. and Jensen, P. (2000). "Latitudinal diversity gradients in the deep sea with special reference to North Atlantic nematodes." Marine Ecology Progress Series **194**: 159-167.

Lawton, J. H., Bignell, D. E., Bloemers, G. F., Eggleton, P. and Hodda, M. E. (1996). "Carbon flux and diversity of nematodes and termites in Cameroon forest soils." Biodiversity and Conservation **7**: 261-273.

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holts, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S. and Watt, A. D. (1998). "Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest." Nature **391**: 72-75.

Lipscomb, D., Platnick, N. and Wheeler, Q. (2003). "The intellectual content of taxonomy: a comment on DNA taxonomy." Trends in Ecology & Evolution **18**: 65-66.

Magurran, A. E. (1988). Ecological Diversity and its Measurement. Chapman and Hall, London.

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

Mallet, J. (1995). "A species definition for the modern synthesis." Trends in Ecology & Evolution **10**: 294-299.

Mallet, J. and Willmott, K. (2003). "Taxonomy: renaissance or Tower of Babel?" Trends in Ecology & Evolution **18**: 57-59.

Markmann, M. (2000). Entwicklung und Anwendung einer 28S rDNA Sequenzdatenbank zur Aufschlüsselung der Artenveilfalt limnischer Meiobenthosfauna im Hinblick auf den Einsatz moderner Chiptechnologie (PhD Thesis). Fakultät für Biologie. Munich, Ludwig Maximilians Universität.

May, R. M. (1975). Patterns of species abundance and diversity. In: Ecology and Evolution of Communities (M. L. Cody and J. M. Diamond, Eds.), pp. 81-120. Harvard University Press, Cambridge, Massachusetts.

May, R. M. (1988). "How many species are there on Earth?" Science **241**: 1441-1449.

Mayden, R. L. (1999). "Consilience and a hierarchy of species concepts: advances towards closure on the species puzzle." Journal of Nematology **31**: 95-116.

Mayr, E. (1963). Animal Species and Evolution. Harvard University Press

McCaig, A. E., Glover, L. A. and Prosser, J. I. (2001). "Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns." Applied and Environmental Microbiology **67**: 4554-4559.

Mikola, J. (1998). "Effects of microbivore species composition and basal resource enrichment on trophic-level biomasses in an experimental microbial-based soil food web." Oecologia **117**(396-403): 396-403.

Mikola, J. and Setälä, H. (1998). "No evidence of Trophic cascades in an experimental microbial-based soil food web." Ecology **79**: 153-164.

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

Myers, N., Mittermeier, R. A., Mittermeier, C. G., G.A.B., d. F. and Kent, J. (2000). "Biodiversity hotspots for conservation priorities." Nature **403**: 853-858.

Nadler, S. A. (1992). "Phylogeny of some Ascaridoid nematodes, inferred from comparison of 18S and 28S rRNA sequences." Molecular Biology & Evolution **9**: 932-944.

Navajas, M. and Boursot, P. (2003). "Nuclear ribosomal DNA monophyly versus mitochondrial DNA polyphyly in two closely related mite species: the influence of life history and molecular drive." Proc. R. Soc. Lond. B. (Suppl.): DOI 10.1098/rsbl.2003.0034.

Nee, S. (2003). "Unveiling prokaryotic diversity." Trends in Ecology & Evolution **18**: 62-63.

Noor, M. A. F. (2002). "Is the biological species concept showing its age?" Trends in Ecology & Evolution **17**: 153-154.

Orr, C. C. and Dickerson, O. J. (1966). "Nematodes in true prairie soils of Kansas." Kansas Academy of Sciences Transactions **69**: 317-334.

OTA (1987). Technologies to Maintain Biological Diversity. US Government Printing Office, Washington DC.

Overgaard Nielsen, C. (1949). "Studies on the soil microfauna II. The soil inhabiting nematodes." Natura Jutlandica **2**: 1-131.

Parkinson, J., Guiliano, D. B. and Blaxter, M. L. (2002). "Making sense of EST sequences by CLOBBing them." BMC Bioinformatics **3**: 31.

Powers, T. O. and Harris, T. S. (1993). "A polymerase chain reaction method for identification of five major *Meloidogyne* species." Journal of Nematology **25**: 1-6.

Powers, T. O., Todd, T. C., Burnell, A. M., Murray, P. C. B., Fleming, C. C., Szalanski, A. L., Adams, B. J. and Harris, T. S. (1997). "The internal transcribed spacer region as a taxonomic marker for nematodes." Journal of Nematology **29**: 441-450.

Preston, F. W. (1948). "The commonness, and rarity, of species." Ecology **29**: 254-283.

Procter, D. L. C. (1990). "Global overview of the functional roles of soil-living nematodes in terrestrial communities and ecosystems." Journal of Nematology **22**: 1-7.

Rappe, M. S., Vergin, K. L. and Giovannoni, S. J. (2000). "Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems." FEMS Microbiol Ecol **33**: 219-232.

Ridley, M. (1996). Evolution, 2nd ed. Blackwell Science

Roca, A. L., Georgiadis, N., Pecon-Slattery, J. and O'Brien, S. J. (2001). "Genetic evidence for two species of African elephant in Africa." Nature **293**: 1473-1477.

Rondon, M. R., August, P. R., Betterman, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J. and Goodman, R. M. (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." Applied and Environmental Microbiology **66**: 2541-2547.

Rosenzweig, M. L. (1995). Species Diversity in Space and Time. Cambridge University Press, Cambridge.

Rosselló-Mora, R. and Amann, R. (2001). "The species concept for prokaryotes." FEMS Microbiology Reviews **25**: 39-67.

Ruess, L., Michelsen, A., Schmidt, I. K., Jonasson, S. and Dighton, J. (1998). "Soil nematode fauna of a subarctic heath: potential nematicidal action of plant leaf extracts." Applied Soil Ecology **7**: 111-124.

Ruess, L., Schmidt, I. K., Michelsen, A. and Jonasson, S. (2001). "Manipulations of a microbial based soil food web at two arctic sites - evidence of species redundancy among the nematode fauna?" Applied Soil Ecology **17**(19-30).

Salles, J. F., De Souza, F. A. and van Elsas, J. D. (2002). "Molecular method to assess the diversity of *Burkholderia* species in environmental samples." Applied and Environmental Microbiology **68**: 1595-1603.

Schulze, E.-D. and Mooney, H. A. (1994). Biodiversity and Ecosystem Function. Springer-Verlag, Berlin.

Schwartz, M. W., Brigham, C. A., Hoeksema, J. D., Lyone, K. G., Mills, M. H. and van Mantgem, P. J. (2000). "Linking biodiversity to ecosystem function: implications for conservation ecology." Oecologia **122**: 297-305.

Seberg, O., Humphries, C. J., Knapp, S., Stevenson, D. W., Petersen, G., Scharff, N. and Andersen, N. M. (2003). "Shortcuts in systematics? A commentary on DNA-based taxonomy." Trends in Ecology & Evolution **18**: 63-65.

Semblat, J. P., Bongiovanni, M., Wajnberg, E., Dalmasso, A., Abad, P. and Castagnone-Sereno, P. (2000). "Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne* spp." Heredity **84**: 81-9.

Semblat, J. P., Wajnberg, E., Dalmasso, A., Abad, P. and Castagnone-Sereno, P. (1998). "High-resolution DNA fingerprinting of parthenogenetic root-knot nematodes using AFLP analysis." Molecular Ecology 7(1): 119-25.

Shannon, C. E. (1948). "A mathematical theory of communication." Bell System Technical Journal **27**: 379-423.

Siddiqi, M. R. (2000). Tylenchida: Parasites of Plants and Insects. CABI Publishing, Oxon.

Simpson, E. H. (1949). "Measurement of diversity." Nature **163**: 688.

Sokal, R. R. and Sneath, P. H. A. (1963). Principles of Numerical Taxonomy. W.H. Freeman & Co., San Francisco.

Southwood, T. R. E. and Henderson, P. A. (2000). Ecological Methods, 3rd ed. Blackwell Science

Stackebrandt, E. and Goebel, B. M. (1994). "Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology." International Journal of Systematic Bacteriology **44**: 846-849.

Stanton, J. M., McNicol, C. D. and 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* 4.0. Sinauer Associates, Sunderland, MA, USA.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996). Phylogenetic Inference. In: Molecular Systematics (D. M. Hillis, C. Moritz and B. K. Mable, Eds.), pp. 407-514. Sinauer Associates, Sunderland, MA, USA.

Szalanski, A. L., Sui, D. D., Harris, T. S. and Powers, T. O. (1997). "Identification of cyst nematodes of agronomic and regulatory concern by PCR-RFLP of ITS1." Journal of Nematology **29**: 253-264.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. and Vogler, A. P. (2003). "A plea for DNA taxonomy." Trends in Ecology & Evolution **18**: 70-74.

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

Theron, J. and Cloete, T. E. (2000). "Molecular techniques for determining microbial diversity and community structure in natural environments." Critical Reviews in Microbiology **26**: 37-57.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Research **25**: 4876-82.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**: 4673-4680.

Thorne, G. (1949). "On the classification of the *Tylenchida*, new order (Nematoda: Phasmidia)." Proc. Helminth. Soc. Wash. **16**: 37-73.

Torsvik, V. and Øvreås, L. (2002). "Microbial diversity and function in soil: from genes to ecosystems." Current Opinion in Microbiology **5**: 240-245.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic Acids Research **23**: 4407-4414.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P. and Trüper, H. G. (1987). "Report of the ad hoc committee on reconciliation of approaches to bacterial systematics." International Journal of Systematic Bacteriology **37**: 463-464.

Werle, E., Schneider, C., Renner, M., Volker, M. and Fiehn, W. (1994). "Convenient single-step, one tube purification of PCR products for direct sequencing." Nucleic Acids Research **22**: 4354-4355.

Wheeler, Q. D. (1999). "Why the phylogenetic species concept? - Elementary." Journal of Nematology **31**: 134.

Wiley, E. O. (1978). "The evolutionary species concept revisited." Systematic Zoology **27**: 17-26.

Wilson, E. O. (2003). "The encyclopedia of life." Trends in Ecology & Evolution **18**: 77-80.

Woese, C. R. (1996). "Phylogenetic trees: Whither Microbiology?" Current Biology **6**: 1060-1063.

Yeates, G. W., Bongers, T., de Goede, R. G. M., Freckman, D. W. and Georgieva, S. S. (1993). "Feeding habits in soil nematode families and genera - an outline for soil ecologists." Journal of Nematology **25**(315-331).

Zarlenga, D. S., Stringfellow, F., Nobary, M. and Lichtenfels, J. R. (1994). "Cloning and characterisation of ribosomal RNA genes from three species of *Haemonchus* (Nematoda:Trichostrongyloidea) and identification of PCR primers for rapid differentiation." Experimental Parasitology **78**: 28-36.

Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. and Hodson, R. E. (2002). "Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20." Applied and Environmental Microbiology **68**: 1576-1584.

Zhou, J., Davey, M. E., Figueras, J. B., Rivkina, E., Gilichinsky, D. and Tiedje, J. M. (1997). "Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA." Microbiology **143**: 3913-9.

# Acknowledgements

I would like to thank my supervisor Mark Blaxter, for showing unending patience and faith in me at all times. His guidance made all of this possible.

I am indebted to both Eyualem Abebe and Artemis Papert, who as Postdocs on the Sourhope project contributed to this work in many ways, and taught me a great deal about nematology and science in general. I also thank every other member of the Blaxter group, past and present, for many kinds of support and encouragement over the years, and for always making the lab an enjoyable place to work: Jennifer Daub, Aziz Aboobaker, Mark Dorris, David Guiliano, Claire Whitton, John Parkinson, Katelyn Fenn, Mark Welsh, Bill Gregory, Marian Thomson, James Wasmuth, Ann Hedley, Alasdair Anthony, Ralf Schmid, Fran Thomas, Habib Maroon and Martin Jones, as well as the many students who have worked with us.

I am grateful to the other members of ICAPB who have provided aid and advice at various times, including Andrew Read and Nick Colegrave (for help with statistics), Sean Nee, Josephine Pemberton, Barbara Wimmer, and Bill Sloan; and to Jill Lovell and Andrew Gillies for their ever-dependable work running the sequencing service.

Sarah Buckland and Graham Burt-Smith, site managers at Sourhope, provided much-appreciated assistance in collecting soil samples. The nematode morphological identification course which I attended at Wageningen University was taught by Tom Bongers, Hanny Van Megen, and Vincent Scholze; they also assisted in identifying some Sourhope nematodes. Paul De Ley, Jaques Vanfleteren, and David Fitch all kindly provided their unpublished SSU sequence data.

I thank my parents for their constant support in all things.

This work was made possible, as part of the Soil Biodiversity Programme, by the financial support of the Natural Environment Research Council (NERC).