



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Imagining Sisyphus happy: DNA barcoding and the unnamed majority

Citation for published version:

Blaxter, M 2016, 'Imagining Sisyphus happy: DNA barcoding and the unnamed majority', *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1702.
<https://doi.org/10.1098/rstb.2015.0329>

Digital Object Identifier (DOI):

[10.1098/rstb.2015.0329](https://doi.org/10.1098/rstb.2015.0329)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Philosophical Transactions of the Royal Society B: Biological Sciences

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Cite this article: Blaxter M. 2016 Imagining Sisyphus happy: DNA barcoding and the unnamed majority. *Phil. Trans. R. Soc. B* **371**: 20150329.

<http://dx.doi.org/10.1098/rstb.2015.0329>

Accepted: 29 April 2016

One contribution of 16 to a theme issue
'From DNA barcodes to biomes'.

Subject Areas:

taxonomy and systematics

Keywords:

metabarcoding, Sisyphus, meiofauna

Author for correspondence:


Mark Blaxter

e-mail: mark.blaxter@ed.ac.uk

Imagining Sisyphus happy: DNA barcoding and the unnamed majority

Mark Blaxter

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK

 MB, 0000-0003-2861-949X

The vast majority of life on the Earth is physically small, and is classifiable as micro- or meiobiota. These organisms are numerically dominant and it is likely that they are also abundantly speciose. By contrast, the vast majority of taxonomic effort has been expended on 'charismatic megabionts': larger organisms where a wealth of morphology has facilitated Linnaean species definition. The hugely successful Linnaean project is unlikely to be extensible to the totality of approximately 10 million species in a reasonable time frame and thus alternative toolkits and methodologies need to be developed. One such toolkit is DNA barcoding, particularly in its metabarcoding or metagenetics mode, where organisms are identified purely by the presence of a diagnostic DNA sequence in samples that are not processed for morphological identification. Building on secure Linnaean foundations, classification of unknown (and unseen) organisms to molecular operational taxonomic units (MOTUs) and deployment of these MOTUs in biodiversity science promises a rewarding resolution to the Sisyphean task of naming all the world's species.

This article is part of the themed issue 'From DNA barcodes to biomes'.

I hope there's an animal somewhere that nobody has ever seen. And I hope nobody ever sees it.

Wendell Berry 'To the unseeable animal' [1].

1. The Linnaean revolution: biology's enduring megaproject

Physics and astronomy are replete with megaprojects, such as the Large Hadron Collider and the Mars missions, that dwarf other modern investments in basic science. But these projects, and the recent huge projects in biology—ENCODE [2], the Human Genome Project [3], the Structural Genomics project [4]—are dwarfed by the longest-running, most successful and most impactful biology megaproject of all: the 263-year old Linnaean project [5,6]. The revolution in biology initiated by Linnaeus in proposing a binomial system for 'naming' groups of plants and animals that were recognizable as distinct natural types has changed the world. These names are a *lingua franca* that can be used to communicate complex concepts and understanding across the globe, helping to organize agriculture, trade and industry, and that is the standard against which we can measure human impact on the planet's ecosystems. The project has been delivered by hundreds of thousands of taxonomists, working lifetimes to describe an estimated 1.2 million species [7]. For every krona, euro, dollar or yen invested in taxonomy, the world economy has likely reaped many-fold returns through identification of disease organisms, invasive species, important biomaterial sources and biomarker taxa.

But how many species of life are there on the Earth? And how close are we to completing the great catalogue? Through careful modelling of species diversity in different groups and encompassing both terrestrial and marine ecosystems, Mora *et al.* [8] estimated that there are 8.7 million species on the Earth, or seven times as many as have been described in the first quarter century of the Linnaean Project.

Their estimate, based on careful modelling of known and predicted species numbers in different phyla and kingdoms, sits near the middle of previous estimates [9,10]. Most of the species (7.7 million) are predicted to be Metazoa, and most of the Metazoa are Arthropoda. The catalogue is not complete. As Mora *et al.* [8] said: ‘In spite of 250 years of taxonomic classification and over 1.2 million species already catalogued in a central database, our results suggest that some 86% of existing species on Earth and 91% of species in the ocean still await description.’

The nature of Linnaean taxonomy is such that naming a new species takes a finite amount of effort, but this effort increases as the Linnaean catalogue becomes more complete. Each proposal for new species must be carefully and precisely placed with reference to existing knowledge. Specific diagnosis has to be offered, and the literature and specimens that underpin the diagnoses of closely related taxa must be minutely examined. Naming the remaining 86% of species is going to become ever more difficult, and even with accelerated publication and digitally available descriptions is likely to require a millennium or more of focused effort. It is thus unlikely that the Linnaean catalogue will ever be complete. This is not to say that it is credible that the Linnaean project should stop, but rather that we cannot expect this Sisyphean task to be completed.

For his continued lack of respect for the order of the gods, Sisyphus was condemned for eternity to roll an immense boulder up a steep hill. Each time he neared the summit, and was about to release the boulder to roll down the other side to freedom, the boulder escaped his grasp and thundered back to the plain. Sisyphus had to return to the foot of the hill to restart his task, forever [Homer *Odyssey* 11.13]. For Linnaean species description, the near-endless task might be cut short by mass extinction.

2. DNA sequencing, barcoding and species description

Molecular data have been included in the primary descriptions of new taxa for 20 years or more. For example, in 1996, the first description [11] of the nematode *Pristionchus pacificus*, a species now used as a genetic, developmental and ecological model organism [12], included nuclear small subunit ribosomal RNA sequence. These sequences that accompany the primary species descriptions have been christened ‘genotypes’ [13]. In bacterial taxonomy, molecular data are necessary for species definition, and soon species descriptions for eukaryotic taxa will also include whole genome data. In a retrofit operation, the taxonomy curators at the National Center for Biotechnology Information’s GenBank database have been assigning the label ‘sequence type’ to sequence accessions where there is clear evidence that they derive from a submitted type culture or specimen [14]. This NCBI effort now includes (April 2016) over 25 000 taxa, with 13 219 Bacteria, 509 Archaea and 11 471 Eukaryota (of which 1141 are Metazoa). The Metazoan species are largely associated with cytochrome oxidase 1 (COI) barcode submissions. A global effort to determine genome sequences for bacterial type strains is underway [15,16]. These data are significantly augmented by the DNA barcode data amassed by the DNA barcoding community and presented in the BOLD System database [17]. BOLD DNA barcode data are associated (in the main) with specimens,

the majority of which in turn have expert morphological species identifications. Thus, sequence-species links are available for over 250 000 Linnaean species of Viridiplantae, Metazoa and Fungi (and approximately 1000 species from four phyla of protists).

3. Neglected animals are largely small

Numerically, the overwhelming majority of individual organisms are microscopic, with major body axes less than 1 mm. While this size class obviously includes (most) Bacteria, Archaea and protists, it also includes the majority of Metazoa, a group famed for its charismatic megataxa. For example, in beach sediments the meiofauna outnumber the indwelling meso- and megafauna by orders of magnitude [18]. However, these massive populations might have a small impact on overall diversity if they include relatively few distinct species. Species number estimates for phyla that are largely meiofaunal, for example, those of Mora *et al.* [8], suggest relatively low total species counts, based on low counts of described species. These estimates can be at odds with focused analyses carried out on smaller branches of the tree of life.

Underestimation of species numbers is particularly prevalent in meiofauna, where lack of easily recognized morphological characters (when the whole specimen is 200 µm long, species specific characters may be at the limit of light microscopy), the incompleteness of early descriptions, the variance and environmental plasticity of form, and the sheer abundance of individuals challenge the working methodologies of systematists. In phylum Nematoda, there are approximately 24 000 described species. Hallan, in a painstaking encyclopaedic effort completed in 2007, catalogued 22 136 species level taxa [19], and nematode taxonomists have been busy since then adding to the catalogue (e.g. 15 new species of *Caenorhabditis* in one paper [20]). However, for Nematoda, the modelling of Mora *et al.* [8] estimated a total species number less than the current described species number. Other authors, using data collected from geographically or ecologically restricted sampling, suggest that true Nematoda species numbers may be anywhere between 1 and 100 million [21–25]. While the upper estimates are deprecated, in particular, owing to a more restricted diversity than expected in deep sea benthos communities [25–27], support for a total of approximately 1 million nematode species is credible. Evidence is also emerging of many cryptic species in Nematoda. For example, the widespread and long-studied *Pellioditis marina* has been shown to be a complex of dozens of reproductively isolated metapopulations (i.e. species) that can be distinguished by DNA sequence and mating tests, but not morphology [28]. A similar finding of extensive cryptic species in large nematodes (greater than 1 cm) [29] suggests that cryptic species may be a common reason for underestimation of species numbers.

There are likely to be many undescribed meiofaunal species, particularly in marine sediments. As marine enoplid nematode specialist Ashleigh Smythe said ‘Marine nematologists get excited when we find a known species’ (personal communication to M.B. 2016). The high estimates of global nematode species numbers [23] are based not on identification of specimens to Linnaean taxa but of sorting of samples into operational taxonomic units (OTUs), and these OTUs are, in turn, likely to conflate cryptic taxa. Identification of nematodes to species is time consuming and requires a high level of

expertise, such that comprehensive nematode species inventories are practically impossible using traditional methodologies [30]. The same set of issues (hyperabundance, hyperdiversity, crypsis and lack of distinguishing morphology) are likely to challenge species identification and diversity assessment in other meiobiotical eukaryotes, such as Tardigrada [31], Oribatida [32] and the many single-celled eukaryote phyla.

4. Metabarcoding accesses meiofauna

It is trivial to set a plankton net or sieve a few tens of grams of sediment and collect hundreds of thousands of specimens of meiofauna. Nematodes achieve astounding numerical abundances, up to a record of $27 \times 10^6 \text{ m}^{-2}$ in an estuary sediment [33]. Like other animals, meiofaunal species abundances follow overdispersion curves, with a few species present in large numbers and most species present rarely. Sifting by eye through a cloud of swimming larvae or thrashing nematodes for rare novelty is a thankless task, and current methodologies subsample from a preserved sample to estimate taxon abundances rather than to enumerate novelty. By repeated sampling, the likely pattern of taxon abundance and diversity can be estimated, but this is costly in researcher time. Advances in sequencing technologies now permit bulk sampling of barcoding genes from a population without the need to individually separate, amplify and sequence each specimen. This approach, pioneered for analysis of the expected 99% of unculturable prokaryotes and microbial eukaryotes [34–36], has been termed metagenetics, environmental barcoding or metabarcoding [37–39]. While the individual specimens are never seen or assessed for Linnaean taxonomy, it is possible to use their sequences as proxies for both their presence (and perhaps abundance) and systematic affinities. The use of PCR in complex mixtures and the inherent errors and biases of the new sequencing technologies mean that careful filtering for errors and artefacts is essential before robust molecular operational taxonomic unit (MOTU) estimation can be carried out [40,41]. With current technologies that involve amplification before sequencing, formation of chimaeric fragments remains a significant issue [41]. The power of new sequencing technologies (generating tens to hundreds of millions of sequences at a time [37]), advanced computational toolkits [42] and ever faster processors now mean that this task is relatively trivial, and as analysis is algorithmic it can be programmed to run automatically [43].

Using this metabarcoding approach a number of groups have started to probe the meiofauna of sediments and soils to ask what is there and to assess whether the ‘extreme’ assertions of meiofaunal species number are likely to be true. Tens of thousands to millions of sequences from a variety of meiofaunal assemblages have now been determined and analysed as MOTU sets [43–46]. These analyses have many notable features. As expected, Nematoda and Arthropoda dominate marine sediments in terms of numbers of reads and of MOTUs. In the same sediments high numbers of Platyhelminthes taxa [44,46] have been found, against expectation. Flatworms are not often observed in abundance in sediments. The disparity between the morphological and barcoding approaches may be because meiofaunal flatworms must be sampled live for morphological identification, as when preserved they are effectively indistinguishable from detritus. However, DNA extraction is agnostic as to appearance and

thus delivers flatworm DNA for amplification and sequencing. Importantly, the DNA barcoding approach was able to affirm the presence of 1000–2500 eukaryote MOTUs per survey [43–46], and classify these robustly by comparison with existing databases of marker genes. The paucity of species-tagged reference sequence data for many phyla sampled in these experiments precludes (in most cases) assignment to species-equivalent in the Linnaean system, but the MOTUs can be robustly assigned to genera or families. This is, in most cases, enough to infer likely life-history characteristics such as mode of feeding.

5. Molecular operational taxonomic units as species

DNA barcode data can be used to define OTU. We originally called these MOTU and suggested that MOTUs were useful proxies for ‘species’ level taxa, and could be used in a similar way to OTUs [47,48]. Importantly, the algorithm used for MOTU definition can be explicit, and largely deterministic. This permits both hypothesis testing and interoperability of MOTU analyses. It is possible to aggregate data across studies and robustly synonymize taxa in different datasets, desynonymize cryptic taxa, co-cluster larval and adult specimens, and identify prey-derived sequences from gut or faecal samples. The BOLD System database [17] now includes a clustering of barcode sequences into MOTUs, called BINs (Barcode Identification Numbers), wherein DNA barcode sequences have been clustered [49] to define groups that have congruent internal divergence and are distinct from groups generated from related sequences. The BIN system has been critical in advancing discussion and discovery of new taxa, and provides a working example of a post-Linnaean taxonomic system. Importantly, as BOLD System and the Barcode of Life project generally are specimen-based, BINs can be linked to specimens, and the Linnaean diagnoses of those specimens. Similarly, the UNITE project aims to deliver a MOTU-based clustering of fungal isolate internal transcribed spacer (ITS) sequence linked, where possible, to named taxa [50]. UNITE currently presents 53 891 ‘fungal species hypotheses’ based on ITS data.

But are MOTUs ‘species’? In the early years of the DNA barcoding field, there was animated debate as to whether COI data, and particularly the partial COI target chosen for DNA barcoding, was sufficient to distinguish animal species, and whether it was an honest marker of species membership. This debate has not been resolved, but it is clear that for most animal taxa the COI barcode is good at distinguishing species, and that exceptions to this general success are both real and interesting, but not fatal to the programme as a whole. In animals, for example, COI generally accumulates substitutions rapidly enough so as to be informative between all but the most recently diverged taxa, and population genetic processes serve to assure that the coalescence of the mitochondrial haplotype is usually before coalescence of the nuclear haplotypes in a species. Issues arise where mitochondria have introgressed between species, sometimes because of other cytoplasmic genetic elements such as intracellular symbionts like *Wolbachia* [51]. In complex species groups, joint analyses of nuclear and mitochondrial markers may be necessary to distinguish species [52]. Alternative markers, such as the nuclear ribosomal RNA subunits, also have promise, particularly as it is possible to derive near-universal

primer sets that amplify all taxa. However, the greater conservation of ribosomal RNA genes that permits universal primer design also limits the specific diagnosis possible with ribosomal RNA sequence. Closely related taxa can be identical in small ribosomal RNA sequence, or so close as to be indistinguishable from sequencing error [48].

Despite these caveats, specimens assigned to MOTUs or BINs can be used in ecological and other surveys just as would specimens assigned to Linnaean taxa [37,43–46]. Individual MOTUs can be assigned abundances and the presence–absence statistics, samples can be compared both within and between studies, and ecological parameters inferred from interaction with abiotic factors and from inferred taxonomic affinities. For meiofaunal surveys, where morphological taxonomy is unable, for operational reasons, to identify OTU to Linnaean species, it can be impossible to cross-compare different ecosystems because it is not possible to synonymize between ‘*Daptonema* sp. 1’ in one publication and the several unallocated *Daptonema* species identified in a second [53].

6. Outlook

Some dung beetles famously collect and roll relatively huge pellets of dung in order to provision their offspring, and are a living model for Sisyphus’s task. Like all biological mechanisms, this activity has no greater purpose than an attempt by dung beetle genes to produce additional copies of themselves in the future. Camus was fascinated by the myth of Sisyphus [54], in particular, in the way it perhaps echoed the

pointlessness of human existence: no matter how hard Sisyphus worked, no matter how often he nearly reached the summit, he always watched his boulder rolling out of control back down the hillside to the plains. But Camus was interested in Sisyphus as he walked back down to restart his task: what was Sisyphus thinking, what did he feel? During the period of this endless return, Camus suggests we must consider Sisyphus happy: happy that the task is still there to be done [54]. Similarly, while we may never complete the cataloguing and systematizing of life on our planet, it is deeply illuminating to attempt to do so. Knowing the limits of diversity, the edges of the puzzle, may be as important as filling in every piece. Especially, as we experience the sixth great extinction it may be a hollow victory to develop a species or MOTU list for a disappearing ecosystem. That said, metagenetics is, I believe, the only way we are going to glimpse the majority of life on the Earth, and, while we may never see the organism we sequence, we will be able to infer its place in phylogeny, impute its biology and consider its roles in driving ecosystem function.

‘La lutte elle-même vers les sommets suffit à remplir un cœur d’homme. Il faut imaginer Sisyphe heureux.’

[The struggle to the summit in itself is enough to fill the heart. We must imagine Sisyphus happy.]

Albert Camus, *Le Mythe de Sisyphe* [54].

Competing interests. I have no competing interests.

Funding. M.B. is funded by NERC and BBSRC.

Acknowledgements. I thank colleagues and students for engaging discussion on barcoding, MOTUs and the nature of species.

References

- Berry W. 1970 *Farming: a hand book*. New York, NY: Harcourt Brace Jovanovich.
- Encode Project Consortium. 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. (doi:10.1038/nature11247)
- Lander ES *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)
- Grabowski M, Niedzialkowska E, Zimmerman MD, Minor W. 2016 The impact of structural genomics: the first quinquennial. *J. Struct. Funct. Genomics* **17**, 1–16. (doi: 10.1007/s10969-016-9201-5)
- Linnaeus C. 1753 *Species plantarum*, 1st edn. Stockholm, Sweden: Laurentius Salvius.
- Linnaeus C. 1758 *Systema naturae*, 10th edn. Stockholm, Sweden: Holmiae Salvius.
- Species 2000 & ITIS. 2010 *Catalogue of life: 2010 annual checklist*. See <http://www.catalogueoflife.org/annual-checklist/2010> (accessed 14 April 2016).
- Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. 2011 How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127. (doi:10.1371/journal.pbio.1001127)
- May RM. 1988 How many species are there on Earth? *Science* **241**, 1441–1449. (doi:10.1126/science.241.4872.1441)
- May RM. 2010 Ecology. Tropical arthropod species, more or less? *Science* **329**, 41–42. (doi:10.1126/science.1191058)
- Sommer RJ, Carta LK, Kim SK, Sternberg PW. 1996 Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda: Neodiplogastriidae). *Fundam. Appl. Nematol.* **19**, 511–521.
- Sommer RJ, McGaughan A. 2013 The nematode *Pristionchus pacificus* as a model system for integrative studies in evolutionary biology. *Mol. Ecol.* **22**, 2380–2393. (doi:10.1111/mec.12286)
- Chakrabarty P. 2010 Genotypes: a concept to help integrate molecular systematics and traditional taxonomy. *Zootaxa* **2632**, 67–68.
- Federhen S. 2015 Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.* **43**(Database issue), D1086–D1098. (doi:10.1093/nar/gku1127)
- Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk H-P. 2014 Genomic encyclopedia of type strains, phase I: the one thousand microbial genomes (KMG-I) project. *Stand. Genomic. Sci.* **9**, 1278–1284. (doi:10.4056/signs.5068949)
- Kyrpides NC *et al.* 2014 Genomic encyclopedia of Bacteria and Archaea: sequencing a myriad of type strains. *PLoS Biol.* **12**, e1001920. (doi:10.1371/journal.pbio.1001920)
- Ratnasingham S, Hebert PD. 2007 BOLD: the barcode of life data system. *Mol. Ecol. Notes* **7**, 355–364. (doi:10.1111/j.1471-8286.2007.01678.x)
- Platonova TA, Gal’tsova VV. 1976 *Nematodes and their role in the meiobenthos*. Leningrad, USSR: Nakua.
- Hallan J. 2007 *Biology catalog: synopsis of the described Nematoda of the world 2007*. See <https://insects.tamu.edu/research/collection/hallan/> (accessed 14 April 2016).
- Felix MA, Braendle C, Cutter AD. 2014 A streamlined system for species diagnosis in *Caenorhabditis* (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. *PLoS ONE* **9**, e94723. (doi:10.1371/journal.pone.0094723)
- Platt HM, Shaw KM, Lamshead PJD. 1984 Nematode species abundance patterns and their use in the detection of environmental perturbations. *Hydrobiologia* **118**, 59–66. (doi:10.1007/BF00031788)
- Platt HM, Lamshead PJD. 1985 Neutral model analysis of patterns of marine benthic species diversity. *Mar. Ecol. Prog. Ser.* **24**, 75–81. (doi:10.3354/meps024075)
- Lamshead PJD. 1993 Recent developments in marine benthic biodiversity research. *Oceanis* **19**, 5–24.
- Platt HM. 1994 Foreword. In *The phylogenetic systematics of free-living nematodes* (ed. S Lorenzen). London, UK: The Ray Society.

25. Lamshead PJD, Boucher G. 2003 Marine nematode deep-sea biodiversity—hyperdiverse or hype? *J. Biogeogr.* **30**, 475–485. (doi:10.1046/j.1365-2699.2003.00843.x)
26. Bik HM, Lamshead PJ, Thomas WK, Lunt DH. 2010 Moving towards a complete molecular framework of the Nematoda: a focus on the Enoplida and early-branching clades. *BMC Evol. Biol.* **10**, 353. (doi:10.1186/1471-2148-10-353)
27. Bik HM, Thomas WK, Lunt DH, Lamshead PJ. 2010 Low endemism, continued deep-shallow interchanges, and evidence for cosmopolitan distributions in free-living marine nematodes (order Enoplida). *BMC Evol. Biol.* **10**, 389. (doi:10.1186/1471-2148-10-389)
28. Derycke S, Backeljau T, Vlaeminck C, Vierstraete A, Vanfleteren J, Vincx M, Moens T. 2006 Seasonal dynamics of population genetic structure in cryptic taxa of the *Pellioditis marina* complex (Nematoda: Rhabditida). *Genetica* **128**, 307–321. (doi:10.1007/s10709-006-6944-0)
29. Derycke S, Sheibani Tezerji R, Rigaux A, Moens T. 2012 Investigating the ecology and evolution of cryptic marine nematode species through quantitative real-time PCR of the ribosomal ITS region. *Mol. Ecol. Resour.* **12**, 607–619. (doi:10.1111/j.1755-0998.2012.03128.x)
30. Lawton JH *et al.* 1998 Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forests. *Nature* **391**, 72–75. (doi:10.1038/34166)
31. Blaxter M, Elsworth B, Daub J. 2004 DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proc. R. Soc. Lond. B* **271**(Suppl. 4), S189–S192. (doi:10.1098/rsbl.2003.0130)
32. Walter DE, Proctor HC. 1999 *Mites: ecology, evolution and behaviour*. Wallingford, UK: CABI Publishing.
33. Warwick RM, Price R. 1979 Ecological and metabolic studies on free-living nematodes from an estuarine mudflat. *Estuar. Coast. Mar. Sci.* **9**, 257–271. (doi:10.1016/0302-3524(79)90039-2)
34. Amaral Zettler LA, Gomez F, Zettler E, Keenan BG, Amils R, Sogin ML. 2002 Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137. (doi:10.1038/417137a)
35. Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A, Sogin ML. 2002 Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc. Natl Acad. Sci. USA* **99**, 7658–7662. (doi:10.1073/pnas.062186399)
36. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006 Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl Acad. Sci. USA* **103**, 12 115–12 120. (doi:10.1073/pnas.0605127103)
37. Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M. 2015 Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PLoS ONE* **10**, e0138432. (doi:10.1371/journal.pone.0138432)
38. Baird DJ, Hajibabaei M. 2012 Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Mol. Ecol.* **21**, 2039–2044. (doi:10.1111/j.1365-294X.2012.05519.x)
39. Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ. 2011 Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* **6**, e17497. (doi:10.1371/journal.pone.0017497)
40. Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT. 2009 Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* **6**, 639–641. (doi:10.1038/nmeth.1361)
41. Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, Power DM, Creer S. 2012 Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Res.* **40**, e66. (doi:10.1093/nar/gks002)
42. Edgar RC. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998. (doi:10.1038/nmeth.2604)
43. Bik HM, Halanych KM, Sharma J, Thomas WK. 2012 Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. *PLoS ONE* **7**, e38550. (doi:10.1371/journal.pone.0038550)
44. Fonseca VG *et al.* 2010 Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1**, 98. (doi:10.1038/ncomms1095)
45. Creer S *et al.* 2010 Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol. Ecol.* **19**, 4–20. (doi:10.1111/j.1365-294X.2009.04473.x)
46. Lallias D *et al.* 2015 Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J.* **9**, 1208–1221. (doi:10.1038/ismej.2014.213)
47. Blaxter ML. 2004 The promise of a DNA taxonomy. *Phil. Trans. R. Soc. Lond. B* **359**, 669–679. (doi:10.1098/rstb.2003.1447)
48. Floyd R, Abebe E, Papert A, Blaxter M. 2002 Molecular barcodes for soil nematode identification. *Mol. Ecol.* **11**, 839–850. (doi:10.1046/j.1365-294X.2002.01485.x)
49. Ratnasingham S, Hebert PD. 2013 A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE* **8**, e66213. (doi:10.1371/journal.pone.0066213)
50. Koljalg U *et al.* 2013 Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271–5277. (doi:10.1111/mec.12481)
51. Jiggins FM, Hurst GD, Schulenburg JH, Majerus ME. 2001 Two male-killing *Wolbachia* strains coexist within a population of the butterfly *Acraea encedon*. *Heredity (Edinb.)* **86**, 161–166. (doi:10.1046/j.1365-2540.2001.00804.x)
52. Nicholls JA, Challis RJ, Mutun S, Stone GN. 2012 Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gallwasps. *Mol. Ecol.* **21**, 4051–4062. (doi:10.1111/j.1365-294X.2012.05683.x)
53. Ferrero TJ, Debenham NJ, Lamshead PJD. 2008 The nematodes of the Thames estuary: assemblage structure and biodiversity, with a test of Attrill's linear model. *Estuar. Coast. Shelf Sci.* **79**, 409–418. (doi:10.1016/j.ecss.2008.04.014)
54. Camus A. 1942 *Le Mythe de Sisyphe*. Paris, France: Éditions Gallimard.