

CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms

Jan Pawlowski^{1*}, Stéphane Audic², Sina Adl³, David Bass⁴, Lassaâd Belbahri⁵, Cédric Berney⁴, Samuel S. Bowser⁶, Ivan Cepicka⁷, Johan Decelle², Micah Dunthorn⁸, Anna Maria Fiore-Donno⁹, Gillian H. Gile¹⁰, Maria Holzmann¹, Regine Jahn¹¹, Miloslav Jirků¹², Patrick J. Keeling¹³, Martin Kostka^{12,14}, Alexander Kudryavtsev^{1,15}, Enrique Lara⁵, Julius Lukeš^{12,14}, David G. Mann¹⁶, Edward A. D. Mitchell⁵, Frank Nitsche¹⁷, Maria Romeralo¹⁸, Gary W. Saunders¹⁹, Alastair G. B. Simpson²⁰, Alexey V. Smirnov¹⁵, John L. Spouge²¹, Rowena F. Stern²², Thorsten Stoeck⁸, Jonas Zimmermann^{11,23}, David Schindel²⁴, Colomban de Vargas^{2*}

1 Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland, **2** Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7144 and Université Pierre et Marie Curie, Paris 6, Station Biologique de Roscoff, France, **3** Department of Soil Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, **4** Department of Life Sciences, Natural History Museum, London, United Kingdom, **5** Laboratory of Soil Biology, University of Neuchâtel, Neuchâtel, Switzerland, **6** Wadsworth Center, New York State Department of Health, Albany, New York, United States of America, **7** Department of Zoology, Charles University in Prague, Prague, Czech Republic, **8** Department of Ecology, University of Kaiserslautern, Kaiserslautern, Germany, **9** Institute of Botany and Landscape Ecology, University of Greifswald, Greifswald, Germany, **10** Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada, **11** Botanischer Garten und Botanischer Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, Germany, **12** Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic, **13** Canadian Institute for Advanced Research, Botany Department, University of British Columbia, Vancouver, British Columbia, Canada, **14** Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, **15** Department of Invertebrate Zoology, St-Petersburg State University, St-Petersburg, Russia, **16** Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, **17** Allgemeine Ökologie, Universität zu Köln, Köln, Germany, **18** Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden, **19** Department of Biology, University of New Brunswick, Fredericton, New Brunswick, Canada, **20** Department of Biology, Life Sciences Centre, Halifax, Nova Scotia, Canada, **21** National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Computational Biology Branch, Bethesda, Maryland, United States of America, **22** Sir Alister Hardy Foundation for Ocean Science, Citadel Hill, Plymouth, United Kingdom, **23** Justus-Liebig-University, Giessen, Germany, **24** Smithsonian Institution, National Museum of Natural History, Washington, DC, United States of America

Animals, plants, and fungi—the three traditional kingdoms of multicellular eukaryotic life—make up almost all of the visible biosphere, and they account for the majority of catalogued species on Earth [1]. The remaining eukaryotes have been assembled for convenience into the *protists*, a group composed of many diverse lineages, single-celled for the most part, that diverged after Archaea and Bacteria evolved but before plants, animals, or fungi appeared on Earth. Given their single-celled nature, discovering and describing new species has been difficult, and many protistan lineages contain a relatively small number of formally described species (Figure 1A), despite the critical importance of several groups as pathogens, environmental quality indicators, and markers of past environmental changes. It would seem natural to apply molecular techniques such as DNA barcoding to the taxonomy of protists to compensate for the lack of diagnostic

morphological features, but this has been hampered by the extreme diversity within the group. The genetic divergence observed between and within major protistan groups greatly exceeds that found in each of the three multicellular kingdoms. No single set of molecular markers has been identified that will work in all lineages, but an international working group is now close to a solution. A universal DNA barcode for protists coupled with group-specific barcodes will enable an explosion of taxonomic research that will catalyze diverse applications.

The undiscovered species diversity among protists may be orders of magnitude greater than previously thought. Surveys of protistan environmental diversity usually based on Sanger sequencing of polymerase chain reaction-amplified 18S rDNA clone libraries revealed an extremely high proportion of sequences that could not be assigned to any described species and in some cases even suggested the presence of several new eukaryotic kingdoms [2,3]. Although some of these sequences have since been shown to be chimeric or long-branch attraction artefacts (caused by

Citation: Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, et al. (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biol* 10(11): e1001419. doi:10.1371/journal.pbio.1001419

Published: November 6, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The initial phase of the Protist Working Group activities presented in this paper were supported by the Consortium for the Barcoding of Life, the European ERA-net program BiodivErsA, under the BioMarkKs project, the French ANR project 09-BLAN-0348 POSEIDON, and the Swiss National Science Foundation 31003A-140766. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Jan.Pawlowski@unige.ch (JP); vargas@sb-roscoff.fr (CdV)

The Community Page is a forum for organizations and societies to highlight their efforts to enhance the dissemination and value of scientific knowledge.

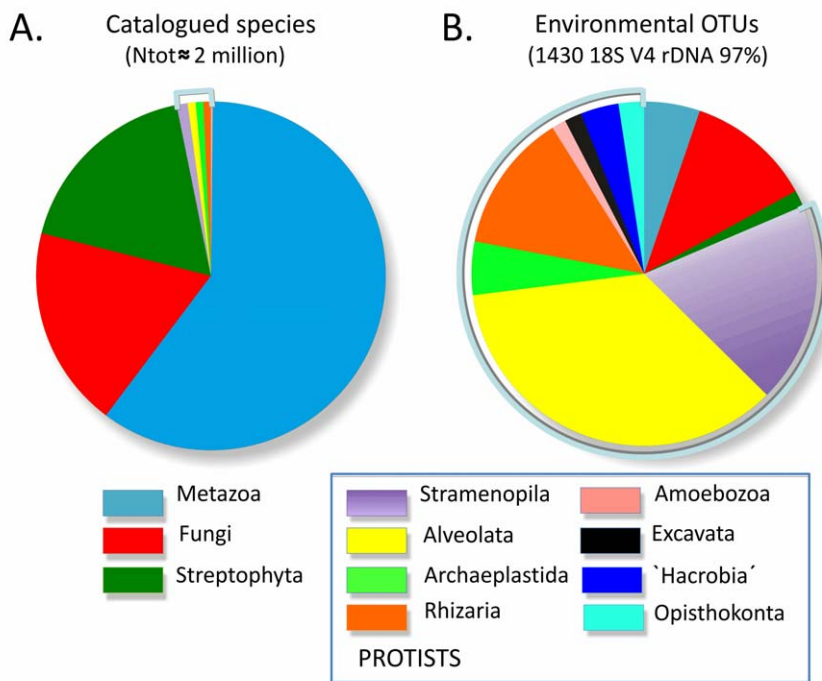


Figure 1. Morphological versus genetic views of total eukaryotic diversity. (A) Relative numbers of described species per eukaryotic supergroup—see Table S1 for a detailed count per division/class. (B) Relative number of V4 18S rDNA Operational Taxonomic Units (97%) per eukaryotic supergroup, based on 59 rDNA clone library surveys of marine, fresh-water, and terrestrial total eukaryotic biodiversity (as listed in [55]). doi:10.1371/journal.pbio.1001419.g001

heterogeneity of evolutionary rates) [4], novel protistan phyla continue to be discovered (e.g., [5,6]). More recently, the growing number of Next Generation Sequencing (NGS) studies of eukaryotic diversity [6–9] has confirmed that the evolutionary and ecological importance of protists is much higher than traditionally thought (Figure 1B) and suggest that the number of protist species may easily exceed one million, although the correct estimation depends on many factors discussed below. The flow of eukaryotic sequence data produced by NGS from environmental DNA extracts is exponentially increasing, but there is currently no way to interpret these sequences in terms of species diversity and ecology.

DNA barcoding is a technique that uses a short standardized DNA region to identify species [10]. Large public reference libraries of DNA barcodes are being developed for animals, plants, and fungi, but there is no general agreement on which region to use for protists. Identifying the standard barcode regions for protists and assembling a reference library are the main objectives of the Protist Working Group (ProWG), initiated by the Consortium for the Barcode of Life (CBOL, <http://www.barcodeoflife.org/>).

The ProWG unites a panel of international experts in protist taxonomy and ecology, with the aim to assess and unify the efforts to identify the barcode regions across all protist lineages, create an integrated plan to finalize the selection, and launch projects that would populate the reference barcode library. Here, we discuss the advantages and limitations of DNA barcodes currently in use and introduce a two-steps barcoding approach to assess protistan biodiversity.

The Unknown Vastness of Protist Richness

The first task of the protist barcoding initiative is to assess species richness in all protistan supergroups. In historically well-studied and biologically well-known taxa, such as higher plants or vertebrates, the number of predicted and described species is relatively similar. The situation is diametrically different for the fungi, for which *catalogued* species comprise ~7% of the predicted species number [1]. It is even worse for protists. The number of catalogued protistan species is very low in comparison to the diversity of animals, plants, and fungi, ranging from ~26,010 excluding marine nonphotosynthetic pro-

tists [1] to ~43,000 [11] and ~74,400 for the novel *ProWG* estimates presented herein (Table S1). Among the seven protistan supergroups (Figure 2A), the most diverse are Stramenopiles, with ~25,000 morphospecies. Over 10,000 described species are also found in Alveolata, Rhizaria, and Archaeplastida (excluding land-plants). Much fewer species have been catalogued for Amoebozoa (~2,400), Excavata (~2,300), and for the unicellular Opisthokonta (~300)—this latter group being dominated by animals and fungi.

The *predicted* richness of protistan species ranges from 1.4×10^5 to 1.6×10^6 [12]. In several groups, the number of predicted species has been arbitrarily estimated to be twice the number of described species [12]. But the true number of species could be several orders of magnitude higher. For example, the Apicomplexa are obligatory parasites, including the malaria agent *Plasmodium* and omnipresent *Toxoplasma*, and thus could reach up to 1.2×10^6 species if we assume a strict specificity to their metazoan hosts. The same argument can be applied to predict extreme species richness in protistan parasites of fishes (e.g., Mesomycetozoa) and plants (e.g., Oomycetes). However, most of these predictions are highly subjective.

Moreover, just like in Bacteria and Archaea [13,14], there is no general agreement on how to define species in protists, and no single species concept can be applied unequivocally to all protistan groups. Molecular studies typically reveal a multitude of genotypes hidden within protist species that have been discovered and described using traditional methods based on morphological criteria (often referred to as “morphospecies”). Reproductive isolation could theoretically be used in differentiating eukaryotic species, but data on the very existence of a sexual phase are very sparse in protists. Mating studies in some “model” systems (e.g., [15,16]) are consistent with the evidence from molecular data that protistan species diversity is greatly underestimated by classical morphological approaches. Overall intraspecific and intragenomic variabilities in environmental protistan populations are still largely unknown, because most genetic studies are carried out on clonal strains maintained in laboratory cultures.

Protist Barcoding: State of the Art

Although the term *DNA barcoding* appeared only recently in the protistological

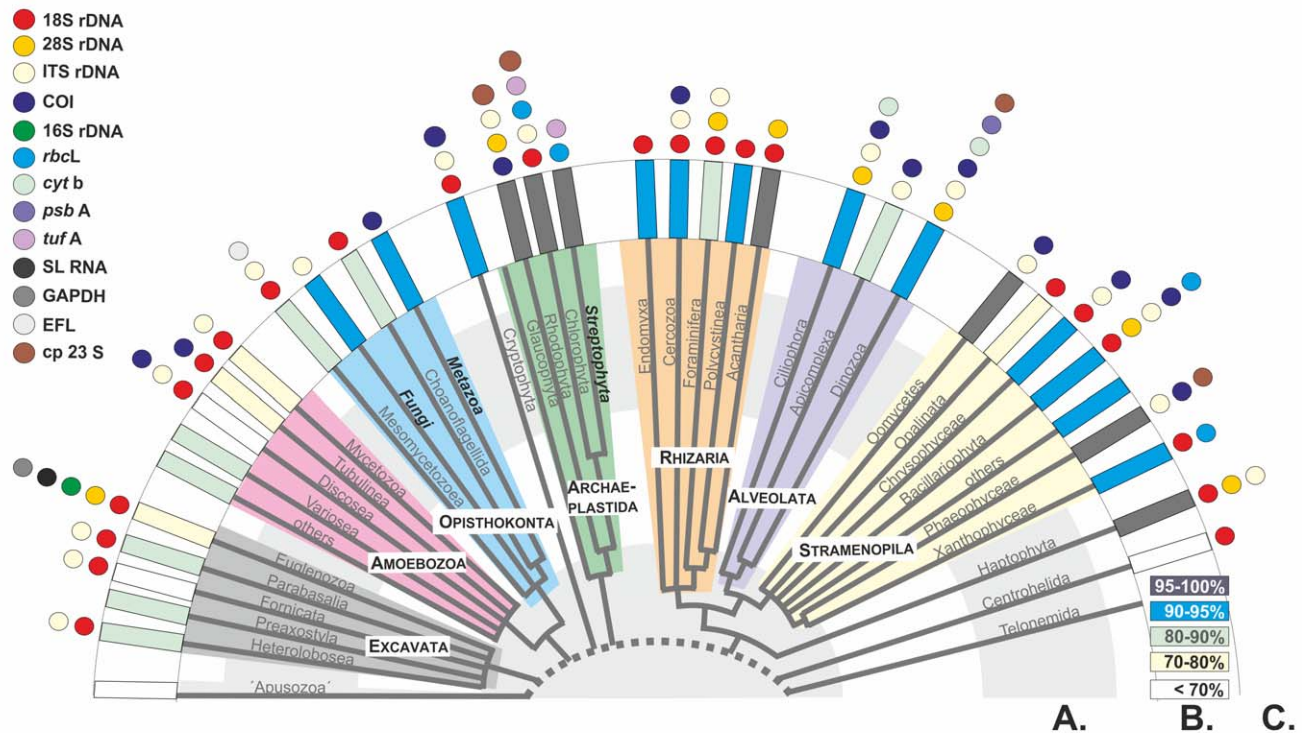


Figure 2. Current state-of-the-art phylogeny and barcode markers for the main protistan lineages. (A) A recent phylogeny of eukaryotic life, after [56]. (B) Mean V4 18S rDNA genetic similarity between all congeneric species within each lineage, available in GenBank. (C) Currently used group-specific barcodes. The dashed line indicates the uncertainty concerning the position of the root in the tree of eukaryotic life. The unresolved relationships between eukaryotic groups are indicated by polytomies. The names of the three multicellular classical “kingdoms” are highlighted. doi:10.1371/journal.pbio.1001419.g002

literature, the identification of protistan taxa using molecular markers has a long history. The most commonly used markers have been parts of the genes coding for ribosomal RNAs, in particular 18S rDNA (e.g., [17]). The advantages of 18S rDNA are many: found in all eukaryotes, it occurs in many copies per genome, allowing genetic work at the individual (single-cell) level; it is highly expressed, permitting molecular ecological investigation at the RNA level; and it includes a mosaic of highly conserved and variable nucleotide sequences allowing combined phylogenetic reconstruction and biota recognition at various taxonomic levels. Different 18S rDNA variable regions have been used in clone libraries and NGS-based environmental surveys [3,7,18]. 18S rDNA barcodes have been shown to effectively distinguish species in some groups, such as foraminifera [19,20] and some diatoms [21], however they are not sufficiently variable to resolve interspecies

relationships in several other taxa (Figure 2B).

Various alternative protistan DNA barcodes have been proposed (Figure 2, Table S2). The D1–D2 and/or D2–D3 regions at the 5' end of 28S rDNA have been positively tested in ciliates [22], haptophytes [23], and acantharians [24] and are also promising for diatoms [25,26]. Ribosomal internal transcribed spacers (ITS1 and/or ITS2 rDNA), which are the main fungal barcodes [27], are also commonly utilized in oomycetes [28], chlorarachniophytes [29], and green algae [30] and have also been suggested for dinoflagellates [31,32] and diatoms [33] with some reserve [34]. The mitochondrial gene coding for cytochrome oxidase 1 (COI), which has been proposed as the universal barcode for animals [10], also allows morpho-species identification in red [35–37] and brown [38,39] algae, dinoflagellates [40], some raphid diatoms [41], Euglyphida [42], lobose naked [43] and

shelled [44] amoebae, coccolithophorid haptophytes [45], and some ciliates [46,47]. Other group-specific barcodes include the large subunit of the ribulose-1,5-biphosphate carboxylase–oxygenase gene (*rbcL*) and the chloroplastic 23S rRNA gene for photosynthetic protists [25,48–50], and Spliced Leader RNA genes for trypanosomatids [51]. Clearly, the choice of group-specific barcodes is often a question of tradition or ease of use, and studies systematically comparing the resolution power of different protistan DNA barcodes are rare [25,42,43,52].

ProWG Objectives and Perspectives

The ultimate objective of the CBOL ProWG is to establish universal criteria for barcode-based species identification in protists. The DNA barcoding approach has several well-known limitations related to the standardization of species identifi-

cation [53,54], and addressing some of the challenges raised by genetic identification of protists will certainly require more fundamental research on protistan speciation. The ProWG will organise workshops and seminars that will provide opportunity to discuss general questions concerning species definition, genetic variations, and applications of DNA barcodes in all protistan groups.

From a practical perspective, the ProWG mission is to establish the genetic standards that will allow recognition of protistan taxa exclusively on the basis of DNA sequence data. Our goal is not to exclude morphological identification but to propose alternative tools that will be more efficient in dealing with the immense protistan biodiversity and more objective and accessible to nonspecialists. In most protistan groups, morphological characters are unreliable for identification at the species level but do provide guides for higher level taxonomic assignments, as well as valuable information about the biology, ecology, and evolution of organisms. Therefore, every protistan reference DNA barcode must be associated with voucher material and/or illustrations providing phenotypic data from the barcoded specimen.

Because of their long, independent, and complex evolutionary histories, protists are so genetically variable that it is virtually impossible to find a single universal DNA barcode suitable for all of them. The ProWG consortium therefore recommends a two-step barcoding approach,

comprising a preliminary identification using a universal eukaryotic barcode, called the pre-barcode, followed by a species-level assignment using a group-specific barcode (Figure 3). In this nested strategy, the ~500 bp variable V4 region of 18S rDNA is proposed as the universal eukaryotic pre-barcode. Group-specific barcodes (Figure 2C) will then have to be defined separately for each major protistan group, based on comparative studies using the CBOL selection criteria, and much of this work is still to be done. Depending on the type of material (isolates and cultures) and whether or not DNA extraction is destructive for the analysed species, the morphological appearance of each barcoded protist will be preserved as microphotographs, fixed cells, or live and/or cryopreserved cultures. This voucher would be deposited in a public collection, just as type specimens are required for new taxa by the nomenclatural codes. Collection details including locality, date, and (as far as possible) habitat characteristics must also be provided, accompanied in parasitic and symbiotic taxa by an accurately identified host voucher or its DNA/tissue sample wherever this is available. Moreover, the extracted DNA must be deposited in a recognized DNA bank or museum collection and cited with a unique identifier to allow checks and further genetic analyses.

Most of these recommendations are already followed where newly described protistan species are based on cultured strains deposited in collections. However, the large majority of protists are currently

uncultivable by known means or not available in culture collections, and genetic data only exist for a very small fraction of described species. Therefore, it is imperative to establish standard barcoding protocols for future protist barcoding projects that will substantially increase the number of collected, described, but uncultivable protists. A combination of novel high-throughput imaging/sorting with newer genetic technologies—including single-amplified-genome methods—opens exciting avenues in protistan metabarcoding. A protist barcoding protocol such as that outlined in Figure 3 will allow collection of the data necessary to set up a representative protist species reference library. The protocols and recommendations concerning protist barcoding will be available at the *ProWG* website (under construction at www.protistbarcoding.org), and a platform dedicated to protist multi-locus barcodes will be accessible at the Barcode of Life Database.

Given the ongoing DNA sequencing revolution, the 21st-century exploration of biodiversity must do more than document the higher macrofaunal and macrofloral branches on the Tree of Life. Amongst other microbes, protists are key but poorly known elements of the ecosystems we see in Nature, including the complex microbiomes hidden within individual plants, animals, and fungi. Ecological models must include protists based on the new knowledge of their species-level diversity that will mostly come from the billions of NGS-generated environmental barcodes. The reference library of standard protistan barcodes will be the Rosetta stone that makes protist diversity less anonymous.

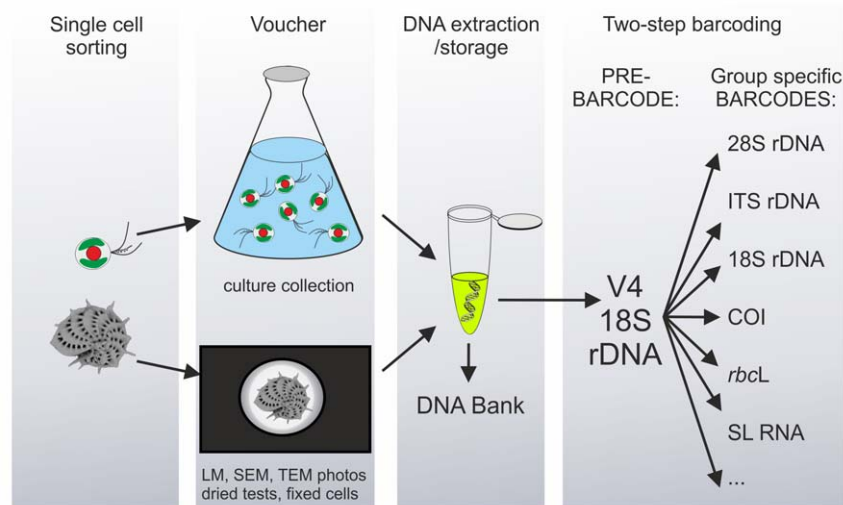


Figure 3. Two-step protist barcoding pipeline. Protistan species, spanning four orders of cell-size magnitude (from $<1 \mu\text{m}$ to $>10,000 \mu\text{m}$), are individually sorted from the environment, phenotyped either directly or after culture growth, DNA extracted, and barcoded using a two-step, nested strategy. doi:10.1371/journal.pbio.1001419.g003

Supporting Information

Table S1 Number of catalogued morphospecies and V4 18S rDNA OTU-97% among the 60 main eukaryotic lineages. (PDF)

Table S2 Group-specific barcodes for selected genera representing all eukaryotic supergroups (in brackets, number of corresponding sequences in the GenBank). NM, nucleomorph origin. Variable regions used in 18S and 28S genes are indicated in some cases. (PDF)

References

- 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
- Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc Natl Acad Sci U S A* 99: 8324–8329.

3. López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409: 603–607.
4. Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic “kingdoms”? Pitfalls and limitations of environmental DNA surveys. *BMC Biol* 2: 13.
5. Kim E, Harrison JW, Sudek S, Jones MD, Wilcox HM, et al. (2011) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc Natl Acad Sci U S A* 108: 1496–1500.
6. Behnke A, Engel M, Christen R, Nebel M, Klein RR, et al. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ Microbiol* 13: 340–349.
7. Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, et al. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* 5: 1344–1356.
8. Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, et al. (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proc Natl Acad Sci U S A* 108: 13177–13182.
9. Logares R, Audic S, Santini S, Pernice MC, de Vargas C, et al. (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J* 6:1823–1833.
10. Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 270: 313–321.
11. Hoef-Emden K, Kupper FC, Andersen RA (2007) Meeting report: Sloan Foundation Workshop to resolve problems relating to the taxonomy of microorganisms and to culture collections arising from the barcoding initiatives; Portland ME, November 6–7, 2006. *Protist* 158: 135–137.
12. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, et al. (2007) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 57: 189–196.
13. Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. *Genome Biology* 7: 116.
14. Caro-Quintero A, Konstantinidis KT (2012) Bacterial species may exist, metagenomics reveal. *Environ Microbiol* 14: 347–355.
15. Coleman AM (1977) Sexual and genetic isolation in the cosmopolitan algal species *Pandorina morum*. *American Journal of Botany* 64: 361–368.
16. Amato A, Kooistra WHCF, Ghiron JHL, Mann DG, Proschold T, et al. (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158: 193–207.
17. Stothard DR, Schroeder-Diedrich JM, Awwad MH, Gast RJ, Ledee DR, et al. (1998) The evolutionary history of the genus *Acanthamoeba* and the identification of eight new 18S rRNA gene sequence types. *J Eukaryot Microbiol* 45: 45–54.
18. Stoock T, Bass D, Nebel M, Christen R, Jones MD, et al. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 19: 21–31.
19. Pawlowski J, Lecroq B (2010) Short rDNA barcodes for species identification in foraminifera. *Journal of Eukaryotic Microbiology* 57: 197–205.
20. de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc Natl Acad Sci U S A* 96: 2864–2868.
21. Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org Divers Evol* 11(3): 173–192.
22. Gentekaki E, Lynn DH (2009) High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the peritrichous ciliate *Carchesium polypinum* in the Grand River basin (North America). *Appl Environ Microbiol* 75: 3187–3195.
23. Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, et al. (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci U S A* 106: 12803–12808.
24. Decelle J, Suzuki N, Mahe F, de Vargas C, Not F (2012) Molecular phylogeny and morphological evolution of the acantharia (radiolaria). *Protist* 163: 435–450.
25. Hamsher SE, Evans KM, Mann DG, Poulickova A, Saunders GW (2011) Barcoding diatoms: exploring alternatives to COI-5P. *Protist* 162: 405–422.
26. Trobajo R, Mann DG, Clavero E, Evans KM, Vanormelingen P, et al. (2010) The use of partial *cox1*, *rbcl* and *LSU* rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* complex (Bacillariophyceae). *Eur J Phycol* 45: 413–425.
27. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A* 109: 6241–6246.
28. Robideau GP, De Cock AW, Coffey MD, Voglmayr H, Brouwer H, et al. (2011) DNA barcoding of oomycetes with cytochrome *c* oxidase subunit I and internal transcribed spacer. *Mol Ecol Resour* 11: 1002–1011.
29. Gile GH, Stern RF, James ER, Keeling PJ (2010) DNA barcoding of Chlorarachniophytes using nucleomorph ITS sequences. *J Phycol* 46: 743–750.
30. Coleman AW (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* 151: 1–9.
31. Litaker RW, Vandersea MW, Kibler SR, Reece KS, Stokes NA, et al. (2007) Recognizing Dinoflagellate species using its rDNA sequences. *J Phycol* 43: 344–355.
32. Stern RF, Andersen R.A., Jameson I, Küpper F.C., Coffroth M.A., et al. (2012) Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. *PLoS One* 7: e42780. doi:10.1371/journal.pone.0042780
33. Moniz MB, Kaczmarek I (2010) Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* 161: 7–34.
34. Mann DG, Sato S, Trobajo R, Vanormelingen P, Souffreau C (2010) DNA barcoding for species identification and discovery in diatoms. *Cryptogamie: Algologie* 31: 557–577.
35. Le Gall L, Saunders G.W. (2010) DNA barcoding is a powerful tool to uncover algal diversity: a case study of the Phylloporaceae (Gigartinales, Rhodophyta) in the Canadian flora. *J Phycol* 46: 374–389.
36. Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos Trans R Soc Lond B Biol Sci* 360: 1879–1888.
37. Saunders GW (2008) A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose *Dilsea-Neodilsea* complex and *Weckisia*. *Botany* 86: 773–789.
38. Kucera E, Saunders G.W. (2008) Assigning morphological variants of *Fucus* (Fucales, Phaeophyceae) in Canadian waters to recognized species using DNA barcoding. *Botany* 86: 1065–1079.
39. McDevitt DC, Saunders GW (2009) On the utility of DNA barcoding for species differentiation among brown macroalgae (Phaeophyceae) including a novel extraction protocol. *Phycological Research* 57: 131–141.
40. Stern RF, Horak A, Andrew RL, Coffroth MA, Andersen RA, et al. (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. *PLoS One* 5: e13991. doi:10.1371/journal.pone.0013991
41. Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom “barcode” genes (*cox1*, *rbcl*, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist* 158: 349–364.
42. Heger TJ, Pawlowski J, Lara E, Leander BS, Todorov M, et al. (2011) Comparing potential COI and SSU rDNA barcodes for assessing the diversity and phylogenetic relationships of cyphodieriid testate amoebae (Rhizaria: Euglyphida). *Protist* 162: 131–141.
43. Nasonova E, Smirnov A, Fahrni J, Pawlowski J (2010) Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist* 161: 102–115.
44. Kosakyan A, Heger TJ, Leander BS, Todorov M, Mitchell EA, et al. (2012) COI barcoding of nebeliid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the hyalospheniidae schultze. *Protist* 163: 415–434.
45. Hagino K, Bendif El M., Young JR, Kogame K, Probert I, et al. (2011) New evidence for morphological and genetic variation in the cosmopolitan coccolithophore *Emiliana huxleyi* (Prymnesiophyceae) from the *cox1b*-ATP4 genes. *J Phycol* 47: 1164–1176.
46. Barth D, Krenk S, Fokin SI, Berendonk TU (2006) Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome C oxidase I sequences. *J Eukaryot Microbiol* 53: 20–25.
47. Chantangsri C, Lynn DH, Brandl MT, Cole JC, Hetrick N, et al. (2007) Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. *Int J Syst Evol Microbiol* 57: 2412–2425.
48. MacGillivray ML, Kaczmarek I (2011) Survey of the efficacy of a short fragment of the *rbcl* gene as a supplemental DNA barcode for diatoms. *J Eukaryot Microbiol* 58: 529–536.
49. Santos S, Gutierrez-Rodriguez C, Coffroth MA (2003) Phylogenetic identification of symbiotic dinoflagellates via length heteroplasmy in domain V of chloroplast large subunit (*cp23S*)—ribosomal DNA sequences. *Mar Biotechnol* (NY) 5(2): 130–140.
50. Sherwood AR, Prestling G.G. (2007) Universal primers amplify a 23S rDNA plastid marker in eukaryotic algal and cyanobacteria. *J Phycol* 43(3): 605–608.
51. Votycka J, Maslov DA, Yurchenko V, Jirku M, Kment P, et al. (2010) Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. *Mol Phylogenet Evol* 54: 243–253.
52. Saunders GW, Kucera H. (2010) An evaluation of *rbcl*, *tufA*, *UPA*, *LSU* and *ITS* as DNA barcode markers for the marine green macroalgae. *Cryptogamie, Algologie* 31: 487–528.
53. Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biol* 2: 1529–1531. doi:10.1371/journal.pbio.0020354
54. Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol Ecol Resour* 12: 377–388.
55. del Campo J, Massana R (2011) Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* 162(3): 435–448.
56. Burki F, Okamoto N, Pombert JF, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* 279:2246–2254.