

UniEuk: Time to Speak a Common Language in Protistology!

Cédric Berney, Andreea Ciuprina, Sara Bender, Juliet Brodie, Virginia Edgcomb, Eunsoo Kim, Jeena Rajan, Laura Wegener Parfrey, Sina Adl, Stéphane Audic, et al.

▶ **To cite this version:**

Cédric Berney, Andreea Ciuprina, Sara Bender, Juliet Brodie, Virginia Edgcomb, et al.. UniEuk: Time to Speak a Common Language in Protistology!. Journal of Eukaryotic Microbiology, 2017, 64 (3), pp.407-411. 10.1111/jeu.12414 . hal-01512649

HAL Id: hal-01512649

https://hal.sorbonne-universite.fr/hal-01512649

Submitted on 24 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

REVIEW ARTICLE

UniEuk: Time to Speak a Common Language in Protistology!

Cédric Berney[a], Andreea Ciuprina[b], Sara Bender[c], Juliet Brodie[d], Virginia Edgcomb[e], Eunsoo Kim[f], Jeena Rajan[g], Laura Wegener Parfrey[h], Sina Adl[i], Stéphane Audic[a], David Bass[d,j], David A. Caron[k], Guy Cochrane[g], Lucas Czech[l], Micah Dunthorn[m], Stefan Geisen[n], Frank Oliver Glöckner[b,o], Frédéric Mahé[p], Christian Quast[o], Jonathan Z. Kaye[c], Alastair G. B. Simpson[q], Alexandros Stamatakis[l,r], Javier del Campo[h], Pelin Yilmaz[o] & Colomban de Vargas[a]

a Sorbonne Universités UPMC Université Paris 06 & CNRS, UMR7144, Station Biologique de Roscoff, Place Georges Teissier, Roscoff 29680, France
b Department of Life Sciences and Chemistry, Jacobs University gGmbH, Bremen D-28759, Germany
c Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, California 94304, USA
d Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom
e Geology and Geophysics Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
f Division of Invertebrate Zoology & Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York 10024, USA
g European Nucleotide Archive, EMBL-EBI, Wellcome Genome Campus, Cambridge CB10 1SD, United Kingdom
h Department of Botany and Zoology, University of British Columbia, 109-2212 Main Mall, Vancouver, BC V6T 1Z4, Canada
i Department of Soil Sciences, College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, SK S7N 5C5, Canada
j Centre for Environment, Fisheries and Aquaculture Science, Barrack Road, Weymouth DT4 8UB, United Kingdom
k Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, California 90089-0371, USA
l Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, Heidelberg D-69118, Germany
m Department of Ecology, University of Kaiserslautern, Kaiserslautern D-67663, Germany
n Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW) & Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 10, Wageningen 6708 PB, The Netherlands
o Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, D-28359, Germany
p CIRAD, UMR LSTM, Montpellier F-34398, France
q Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, NS B3H 4R2, Canada
r Karlsruhe Institute of Technology, Institute for Theoretical Informatics, Postfach 6980, Karlsruhe 76128, Germany

Keywords
Community expertise; diversity; eukaryotes; *EukBank*; *EukMap*; *EukRef*; taxonomy.

Correspondence
P. Yilmaz, Microbial Genomics and Bioinformatics Research Group – Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen 28359, Germany
Telephone number: +49-421-2028-971;
FAX number: +49-421-2028-580;
e-mail: pyilmaz@mpi-bremen.de
and
C. de Vargas, Station Biologique de Roscoff, UMR7144 – Sorbonne Universités UPMC Université Paris 06 & CNRS, Place Georges Teissier, Roscoff 29680, France
Telephone number: +33-2-98-29-25-28;
FAX number: +33-2-98-29-23-24;
e-mail: c2vargas@gmail.com

Received: 10 March 2017; accepted March 13, 2017.

doi:10.1111/jeu.12414

ABSTRACT

Universal taxonomic frameworks have been critical tools to structure the fields of botany, zoology, mycology, and bacteriology as well as their large research communities. Animals, plants, and fungi have relatively solid, stable morpho-taxonomies built over the last three centuries, while bacteria have been classified for the last three decades under a coherent molecular taxonomic framework. By contrast, no such common language exists for microbial eukaryotes, even though environmental '-omics' surveys suggest that protists make up most of the organismal and genetic complexity of our planet's ecosystems! With the current deluge of eukaryotic meta-omics data, we urgently need to build up a universal eukaryotic taxonomy bridging the protist *-omics* age to the fragile, centuries-old body of classical knowledge that has effectively linked protist taxa to morphological, physiological, and ecological information. *UniEuk* is an open, inclusive, community-based and expert-driven international initiative to build a flexible, adaptive universal taxonomic framework for eukaryotes. It unites three complementary modules, *EukRef*, *EukBank*, and *EukMap*, which use phylogenetic markers, environmental metabarcoding surveys, and expert knowledge to inform the taxonomic framework. The *UniEuk* taxonomy is directly implemented in the European Nucleotide Archive at EMBL-EBI, ensuring its broad use and long-term preservation as a reference taxonomy for eukaryotes.

© 2017 The Author(s) Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

THE bewildering organismal and functional complexity of microbial eukaryotes has long fascinated protistologists but exceeded the capacity of this research community to comprehensively study it. Lacking the critical mass for a strong scientific discipline, protistologists remain largely divided into various sub-communities (protozoology versus phycology, aquatic versus terrestrial systems, fossil versus extant organisms, etc.), many of which are adjuncts of other larger fields that may speak different technical languages or use different taxonomic systems. Today, environmental '-omics' surveys make it possible to explore the boundaries of the total biotic diversity in ecosystems, from viruses to animals (Bork et al. 2015). These studies indicate that microbial eukaryotes comprise a huge amount of the organismal and genetic complexity of our planet's biomes, potentially even the majority (e.g. Mahé et al. 2017; de Vargas et al. 2015). This discovery increases the challenge of studying the full complexity of protists, but, at the same time, provides an exceptional opportunity to unite and strengthen the field of modern protistology. For this to occur, it is paramount to construct a universal taxonomy for eukaryotes, a common language that will help unify the field and connect the deluge of new molecular-genetic datasets with each other and with the centuries of accumulated morphological, physiological, life history, and ecological information on these organisms.

UniEuk (http://www.unieuk.org/) is an open, inclusive, community-based and expert-driven international initiative to build a flexible, adaptive universal taxonomic framework for eukaryotes that represents the views of the research community. The effort is focused primarily on protists, and can also incorporate existing taxonomic systems for animals, plants, and fungi. Organism-based and informed by phylogeny, the *UniEuk* framework integrates expert knowledge about morphology and ecology with key molecular information from phylogenetics and environmental '-omics' surveys to capture our total current knowledge on eukaryotic diversity, evolution, and ecology. Its power resides primarily on bottom-up community efforts organized around all protist clades to capture collective knowledge on eukaryotic diversity, with validation of the taxonomic framework by an extensive network of experts. The system's broad use and preservation will be ensured by a direct implementation of the *UniEuk* taxonomy into the European Nucleotide Archive (ENA) at EMBL-EBI (http://www.ebi.ac.uk/ena), with the long-term goal of becoming the reference taxonomy in all INSDC genetic data repositories.

UniEuk was launched in May 2016 with initial funding from the Gordon and Betty Moore Foundation (http://www.moore.org) and the International Society of Protistologists. During the first year, the project's taxonomy and database coordinators and members of the *UniEuk* Steering, Advisory, and Technical Committees (http://www.unieuk.org/people/) designed the three main complementary and interconnected modules for direct community interaction—*EukRef*, *EukBank*, and *EukMap*—that together constitute the core of the *UniEuk* system and are necessary to build the universal taxonomic framework (below and Fig. 1). They also established a baseline version of the *UniEuk* taxonomic framework, starting from existing systems (e.g. Adl et al. 2012), and integrating the most up-to-date information from phylogenomic evidence (e.g. Burki et al. 2016). Lastly, they proposed a set of guidelines for naming environmental genetic lineages prior to their morphological characterization.

(1) *EukRef* (Fig. 1B): The *EukRef* module allows integration into the *UniEuk* system of all preexisting phylogenetic information on eukaryotic diversity derived from Sanger-sequenced DNA markers of described taxa and environmental clones (beginning with 18S rDNA sequences longer than 500 bp). *EukRef* uses a standardized, open-source bioinformatics pipeline to generate homogenous, high-quality curation of 18S rDNA sequences available in the INSDC databases. *EukRef* outputs include, on a lineage-specific basis, taxonomically curated 18S rDNA sequences with corresponding sequence alignments and phylogenetic trees. In addition to being a direct source of information for the *UniEuk* taxonomic framework, *EukRef* outputs represent standalone community resources shared through partner 18S rDNA reference databases SILVA (Quast et al. 2013) and PR^2 (Guillou et al. 2013). *EukRef* has thus far largely engaged PhD students and postdocs who learned how to use the pipeline during multiday workshops, progressively expanding to include all eukaryotic clades. Information on *EukRef* and on past and future workshops can be found at http://www.eukref.org/.

(2) *EukBank* (Fig. 1C): The *EukBank* module allows integration into the *UniEuk* system of the enormous and largely nameless genetic information on eukaryotic diversity obtained from high-throughput metabarcoding (HTM) surveys of the Earth's ecosystems. Combining an ultra-fast algorithm generating stable clusters of amplicons (Mahé et al. 2015) and state-of-the-art methods of phylogenetic placement (Berger et al. 2011), *EukBank* will absorb and reduce the complexity of eukaryotic HTM datasets, and analyze them phylogenetically. Datasets from all planetary biomes will be incorporated, starting with the V4 18S rDNA marker (Pawlowski et al. 2012). *EukBank* is centralized at ENA, providing the community with a protocol and direct assistance for the submission of HTM datasets and their critical metadata to the repository. *EukBank* will allow monitoring of total eukaryotic diversity (e.g. saturation, phylogeny) across biomes, as well as identification and preliminary naming of novel eukaryotic lineages of ecological and/or phylogenetic relevance. These will be integrated into the *UniEuk* taxonomic framework, thus highlighting parts of the tree of eukaryotic life warranting deeper investigation. *EukBank* is aimed at all scientists who have generated eukaryotic HTM datasets and are interested in discovering how these contribute to a growing global perspective on eukaryotic diversity.

(3) *EukMap* (Fig. 1D): The *EukMap* module allows the community to directly interact with and inform the

© 2017 The Author(s) Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

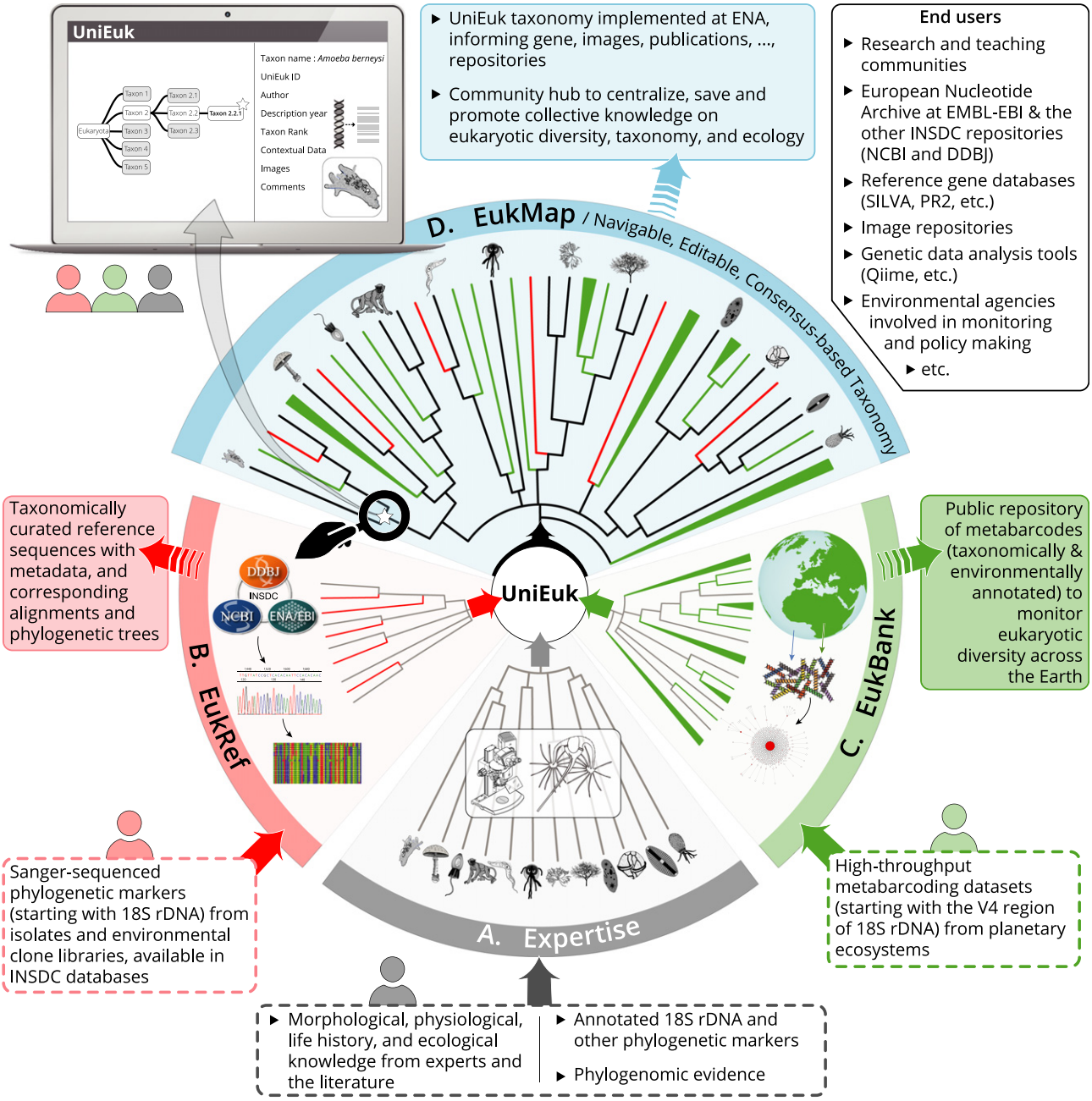


Figure 1 The *UniEuk* workflow. Bottom-up, community-based information on eukaryotic biodiversity from (A) classical knowledge, (B) phylogenetic diversity, and (C) environmental '-omics' surveys, converge and synergize through the *UniEuk* modules to inform the navigable and editable, consensus-based taxonomic framework (D). Dotted and colored frames indicate input and output information, respectively. Line drawings of eukaryotes adapted with permission from https://genev.unige.ch/system/pawlowski/lab/tree.png.

growing universal taxonomic framework. *EukMap* is a user-friendly representation of the *UniEuk* taxonomic framework, publicly navigable, where each node/taxon is associated with standardized features (name, contextual data, links to representative pictures, etc.). *EukMap* will integrate curated genetic information from *EukRef* and *EukBank*, and represents a community hub to centralize, safeguard, and promote our current global knowledge on eukaryotic diversity, taxonomy, and ecology. The output of *EukMap* (the actual *UniEuk* taxonomy) will be directly applied to ENA at EMBL-EBI, with regular versioning. It will provide continuous feedback to the other *UniEuk* modules and partner reference gene databases for optimized and standardized taxonomic annotation of environmental sequence data. It will also be useable as a stand-alone summary of the collective knowledge on

© 2017 The Author(s) Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Box 1: UniEuk timeline and ICOP 2017 announcement
The three main *UniEuk* modules are in the process of implementation. The full *UniEuk* system will be demonstrated to the community at the next joint International Congress of Protistology (ICOP 2017) and Annual Meeting of the International Society of Protistologists in Prague, Czech Republic, July 30 to August 4, 2017 (http://www.icop2017.org/). We encourage all congress participants to attend the *UniEuk* session that will take place the afternoon of Thursday August 3, 2017. After this, the project will be ready to interact with those of you not yet involved: especially experts in protist biodiversity and taxonomy! Visit our website to preregister and be kept updated when the various functionalities of the system become available (http://www.unieuk.org/register/).

eukaryotic diversity and evolutionary history for scientific, educational, or public outreach purposes. *EukMap* is aimed at all researchers, from students to professors, with expertise on eukaryotic taxonomy, ecology, and evolution. The taxonomic framework as a whole is freely navigable for all visitors. Once registered into the *UniEuk* system, community members will be able to propose changes and addition of missing taxa or contextual information (including images) at any taxonomic level, and participate in group-specific discussions to reach agreement on which changes should be adopted in official releases of the *UniEuk* taxonomy. Lead taxonomy experts and the project taxonomy and database coordinators will be in charge of moderating these discussions and implementing decisions.

Further information on *UniEuk*'s vision, goals, and organization can be found at http://www.unieuk.org/, together with use-case scenarios and FAQs, a news section, and a registration page. We encourage all scientists with expertise in protist taxonomy, ecology, or evolution to join the growing *UniEuk* community (Box 1). The deluge of emerging eukaryotic genetic data makes this the right time to build a common, eco-morpho-genetic and organism-centered language for protistology. With its complementary, yet independent modules (all with stand-alone outputs benefitting end-users in the protistology community and beyond; see Fig. 1), we believe that *UniEuk* will generate the necessary momentum to address this challenge. Crucially, the direct application of the *UniEuk* taxonomic framework to all INSDC sequence records via the ENA node at EMBL-EBI means that your contributions will have a direct, immediate impact on global eukaryotic research efforts worldwide. Together, let us make the 21st century the age of protistology!

ACKNOWLEDGMENTS

This study was funded by the Gordon and Betty Moore Foundation through grant GBMF5257, the International Society of Protistologists, and the French Government "Investissements d'Avenir" program OCEANOMICS (ANR-11-BTBR- 0008). We sincerely thank all members of the *UniEuk* committees for their critical input, feedback, and contributions—in the Advisory Council: Sandra Baldauf, Sonya Dyhrman, Laura Katz, Connie Lovejoy, Alexandra Worden, John Archibald, David Bass, David Caron, Patrick Keeling, and Jan Pawlowski; in the Scientific and Technical Advisory Board: Linda Amaral Zettler, Claire Gachon, Laure Guillou, Line Le Gall, Laura Wegener Parfrey, Stéphane Audic, Matthew Brown, Micah Dunthorn, Enrique Lara, Frédéric Mahé, Ramon Massana, and Alexandros Stamatakis; and in the Steering Committee: Juliet Brodie, Virginia Edgcomb, Eunsoo Kim, Pelin Yilmaz, Sina Adl, Guy Cochrane, Colomban de Vargas, Javier del Campo, Stefan Geisen, Frank Oliver Glöckner, and Alastair Simpson.

LITERATURE CITED

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Wegener Parfrey, L., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A. & Spiegel, F. W. 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.*, 59:429–493.

Berger, S. A., Krompass, D. & Stamatakis, A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.*, 60:291–302.

Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E. & Wincker, P. 2015. *Tara* Oceans studies plankton at planetary scale. *Science*, 348:873.

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B. Q., Radaykina, L. V., Smirnov, A., Mylnikov, A. P. & Keeling, P. J. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B*, 283:20152802.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. 2013. The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41:D597–D604.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., Siemensmeyer, T., Trautmann, I., Romac, S., Berney, C., Kozlov, A. M., Mitchell, E. A. D., Seppey, C. V. W., Egge, E., Wirth, R., Trueba, G. & Dunthorn, M. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.*, 1:0091.

Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *Peer J*, 3:e1420.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D. G., Mitchell, E. A. D., Nitsche, F., Romeralo, M.,

© 2017 The Author(s) Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Saunders, G. W., Simpson, A. G. B., Smirnov, A. V., Spouge, J. L., Stern, R. F., Stoeck, T., Zimmermann, J., Schindel, D. & de Vargas, C. 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.*, 10:e1001419.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acid Res.*, 41:D590–D596.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J. M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans Coordinators, Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P. & Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605.

© 2017 The Author(s) Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists