Current Biology
Report

Morphological Identification and Single-Cell Genomics of Marine Diplonemids

Ryan M.R. Gawryluk,[1,*] Javier del Campo,[1] Noriko Okamoto,[1] Jürgen F.H. Strassert,[1] Julius Lukeš,[2,3] Thomas A. Richards,[3,4] Alexandra Z. Worden,[3,5] Alyson E. Santoro,[3,6] and Patrick J. Keeling[1,3,7,*]
[1]Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada
[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences and Faculty of Sciences, University of South Bohemia, 370 05 České Budějovice, Czech Republic
[3]Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, ON M5G 1Z8, Canada
[4]Department of Biosciences, University of Exeter, Geoffrey Pope Building, Exeter EX4 4QD, UK
[5]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, USA
[6]Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA 93106, USA
[7]Lead Contact
*Correspondence: ryan.gawryluk@gmail.com (R.M.R.G.), pkeeling@mail.ubc.ca (P.J.K.)
http://dx.doi.org/10.1016/j.cub.2016.09.013

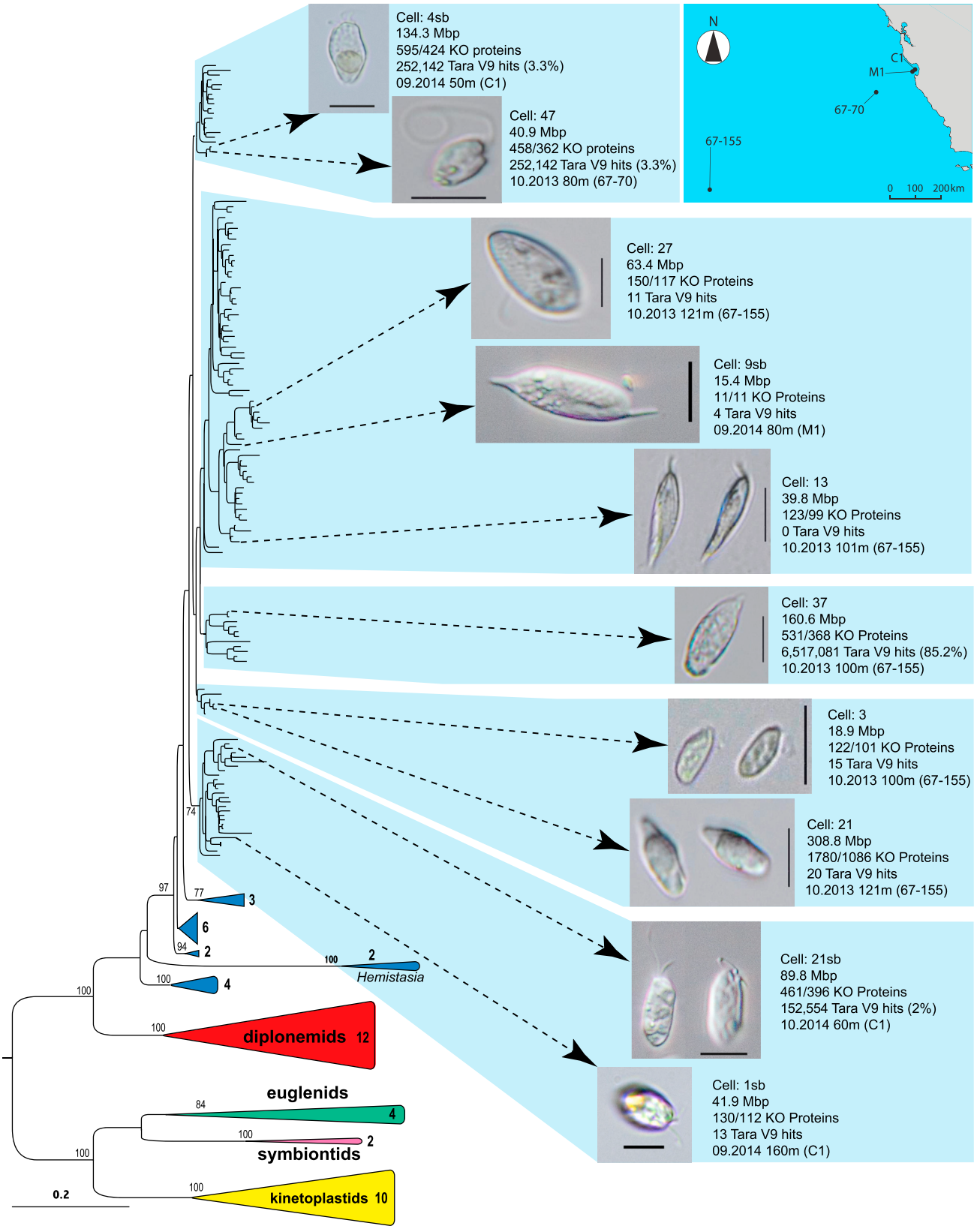
SUMMARY

Recent global surveys of marine biodiversity have revealed that a group of organisms known as "marine diplonemids" constitutes one of the most abundant and diverse planktonic lineages [1]. Though discovered over a decade ago [2, 3], their potential importance was unrecognized, and our knowledge remains restricted to a single gene amplified from environmental DNA, the 18S rRNA gene (small subunit [SSU]). Here, we use single-cell genomics (SCG) and microscopy to characterize ten marine diplonemids, isolated from a range of depths in the eastern North Pacific Ocean. Phylogenetic analysis confirms that the isolates reflect the entire range of marine diplonemid diversity, and comparisons to environmental SSU surveys show that sequences from the isolates range from rare to superabundant, including the single most common marine diplonemid known. SCG generated a total of ~915 Mbp of assembled sequence across all ten cells and ~4,000 protein-coding genes with homologs in the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology database, distributed across categories expected for heterotrophic protists. Models of highly conserved genes indicate a high density of non-canonical introns, lacking conventional GT-AG splice sites. Mapping metagenomic datasets [4] to SCG assemblies reveals virtually no overlap, suggesting that nuclear genomic diversity is too great for representative SCG data to provide meaningful phylogenetic context to metagenomic datasets. This work provides an entry point to the future identification, isolation, and cultivation of these elusive yet ecologically important cells. The high density of nonconventional introns, however, also portends difficulty in generating accurate gene models and highlights the need for the establishment of stable cultures and transcriptomic analyses.

RESULTS AND DISCUSSION

Isolation and Identification of Marine Diplonemids

The diversity and ecological importance of heterotrophic protists (microbial eukaryotes) has been increasingly recognized in recent years [5], but they remain one of the more poorly studied fractions of most ecosystems, largely due to technical challenges. This is exemplified by the case of marine diplonemids: this group of marine flagellates—members of the phylum Euglenozoa—was recorded over a decade ago [2, 3] and recently shown to be among the most abundant and diverse lineages of marine heterotrophs in a massive global environmental survey of sunlit marine waters [1]. Nevertheless, the lineage is only known from a single environmental gene, the small subunit rRNA (SSU rRNA): we know nothing of their morphology; behavior; or basic biology. To characterize a range of heterotrophic marine protists, we carried out manual single-cell isolations from a range of depths (50–160 m) at several stations off of the California (USA) coast, ranging from near coastal to ~800 km offshore, in 2013 and 2014 (Table S1; see Supplemental Experimental Procedures). Because there are no morphological data upon which to base identification, we isolated and photographed 92 colorless flagellates of 10–30 μm with a range of morphologies consistent with expectations for eukaryotic predators (see below). Multiple displacement amplification (MDA) was performed on all isolated cells, and the SSU rRNA was sequenced from all MDA samples (Supplemental Experimental Procedures). Overall, sequence data allowed us to establish the identity of 40 cells, and marine diplonemids accounted for 25%, consistent with the notion that they are abundant planktonic eukaryotes [1]. Other identified lineages include marine alveolates, telonemids, heliozoans, cercozoans, acantharians, stramenopiles, and katablepharids. Given that the relative abundance of diplonemids increases with depth [6], they may be even easier to isolate in deeper waters, which is another possible reason why they have been largely overlooked.



(legend on next page)

"Classical" diplonemids (e.g., *Diplonema* and *Rhynchopus*) are colorless, oblong or elliptical cells, approximately 20 μm in length, with an apical papillum and two flagella that emerge from the subapical depression [7]. The marine diplonemids identified here are diverse in size (~7–25 μm) and shape, but several share specific aspects of morphology expected of a phagotrophic diplonemid. In particular, all were observed to be flagellate (see cells 47, 27, and 21sb in Figure 1), and distinctive apical papillae were observed (Figure 1, cells 9sb and 13).

Phylogenetic Analysis of Marine Diplonemids

Maximum likelihood (ML) phylogenetic analysis of the SSU rRNA gene recovered a strongly supported division within diplonemids, with one clade comprising the cultured classical diplonemids and the second major group comprised of marine planktonic environmental clones. The only cultured representative of the non-classical diplonemids is *Hemistasia*, which is part of a small group of marine environmental clones that is strongly supported to fall outside the major clade of abundant marine diplonemids (Figure 1), so it is unclear how well *Hemistasia* represents the majority of the marine diplonemid diversity. The SSU tree fails to resolve most of the internal branching order of the marine clade, with a few exceptions, such as the clade, including cells 1sb and 21sb. But the isolated strains are nevertheless interspersed across the tree of environmental sequences (Figure 1), suggesting they span the diversity of the group and probably better represent the marine diplonemids than *Hemistasia*. Overall, the phylogeny strongly supports the existence of genetically distinct diplonemid lineages unified by ecology: classical diplonemids are predominantly benthic, whereas "marine" diplonemids are planktonic. Accordingly, it may be more accurate to refer to the classical diplonemids as "benthic" diplonemids and marine diplonemids as "planktonic" diplonemids.

Single-Cell Genomic Analysis of Marine Diplonemids

The size of individual diplonemid single-cell genomics (SCG) assemblies after decontamination (Supplemental Experimental Procedures) ranged from 16 to 303 Mbp, totaling ~915 Mbp between the ten isolates (Figures 1 and S1; Table S2). The number of putative protein-coding genes, inferred by querying the KEGG Automatic Annotation Server [8], was similarly variable (11–1,780; Table S3), yielding ~4,000 in total, including redundant KEGG hits for each SCG assembly. Genes encoding components of numerous biochemical pathways and cellular structures were identified, including subunits of the mitochondrial electron transport chain, histones, paraflagellar rods, and core meiotic machinery (Table S3). However, assessment of genome completeness with BUSCO [9] indicated that the assemblies are very incomplete: the most-complete assembly is missing 91% of conserved single copy orthologs. The overall gene density points to marine diplonemids possessing bloated, gene-sparse, nuclear genomes.

Little is known about the structure of diplonemid nuclear genes in general, and nothing is known about the marine clade. Like other euglenozoans, nuclear gene expression in classical diplonemids requires spliceosome-dependent *trans* splicing of mRNAs [10]. We identified genes encoding spliced-leader (SL) RNAs in all of the marine diplonemid genomes, and at least one from each species encoded SL-RNA that was identical to the 39-nt *Diplonema papillatum* homolog (and some that were only identical over the last ~30 nt). Using primers based on the SL-RNA gene will therefore be an effective way to enrich for mRNAs from a phylogenetically broad assortment of marine diplonemid mRNAs in environmental or single-cell transcriptomic studies.

Little else is known about the structure of diplonemid genes that might help establish the gene models that would be essential to interpret genomic or environmental metagenomic data. Spliceosomal introns, for example, are sparsely known in classical diplonemids [11] but are very rare in kinetoplastids, the closest relatives of diplonemids. Their more-distant euglenoid relatives contain both canonical spliceosomal introns and a small proportion of poorly characterized non-canonical introns that lack GT-AG splice boundaries [12–15]. We carefully examined the most highly conserved genes to allow the identification of introns in the SCG assemblies, as transcriptome data are unavailable, and found a high density of non-canonical introns. These introns lack GT-AG splice boundaries (Figure 2A), and indeed, there is no clear conservation of splice site sequences at all, though there is a slight overrepresentation of both G and C at both splice sites (Figure 2B). Instead, introns frequently have extensive base pairing potential between 5′ and 3′ splice sites and frequently short (3–6 bp) direct repeats, which are partly exonic and partly intronic (Figure 2A). Because the splice site boundaries are repeats, precise determination of intron splice sites is not possible, because it is not clear which of the repeats is exonic; this issue would persist even if orthologous transcriptome data were available. The length of the non-canonical introns ranges from 38 to 8,039 nt (average 409 nt; n = 122), and in some genes, very short exons were observed (e.g., a 6-bp exon is inferred in an alpha-tubulin gene from cell 21).

The evolutionary origin of non-canonical introns in euglenoids is unknown, and the splicing mechanism is thought to be spliceosome independent, as the 5′ splice site is not complementary to the U1 small nuclear RNA (snRNA) [15]. We also identified candidate U1 snRNA genes in six of ten SCG assemblies. U1 is required for binding the 5′ splice site in *cis*-spliced spliceosomal introns but also binds the spliced leader in SL-based *trans* splicing. Candidate U1 species in marine diplonemids are predicted to bind 5′ splice site motifs similar to those found in

Figure 1. Morphology and Evolution of Marine Diplonemids

A euglenozoan-rooted schematic based on maximum likelihood phylogenetic analysis of the diplonemid 18S rRNA gene is shown (1,802 sites; 175 operational taxonomic units [OTUs]). Bootstrap values less than 70 are not shown. Bold numbers refer to the number of taxa present in collapsed clades. Dotted lines connect the phylogenetic placement of each cell's 18S rRNA gene to a photo of that cell. SCG assembly size, number of redundant/non-redundant KO homologs, number of hits to Tara Oceans V9 data (and % of total Discoba hits), collection date, station, and depth are presented beside each photo (for complete collection data, see Table S1). The scale bars represent 10 μm. "Classical" diplonemids (such as *Diplonema* and *Rhynchopus*) group within the red triangle labeled "diplonemids"; *Hemistasia* falls within a small group at the base of the planktonic group. A map of collection sites is in the top right corner. See also Figure S1 and Tables S1, S2, and S3.

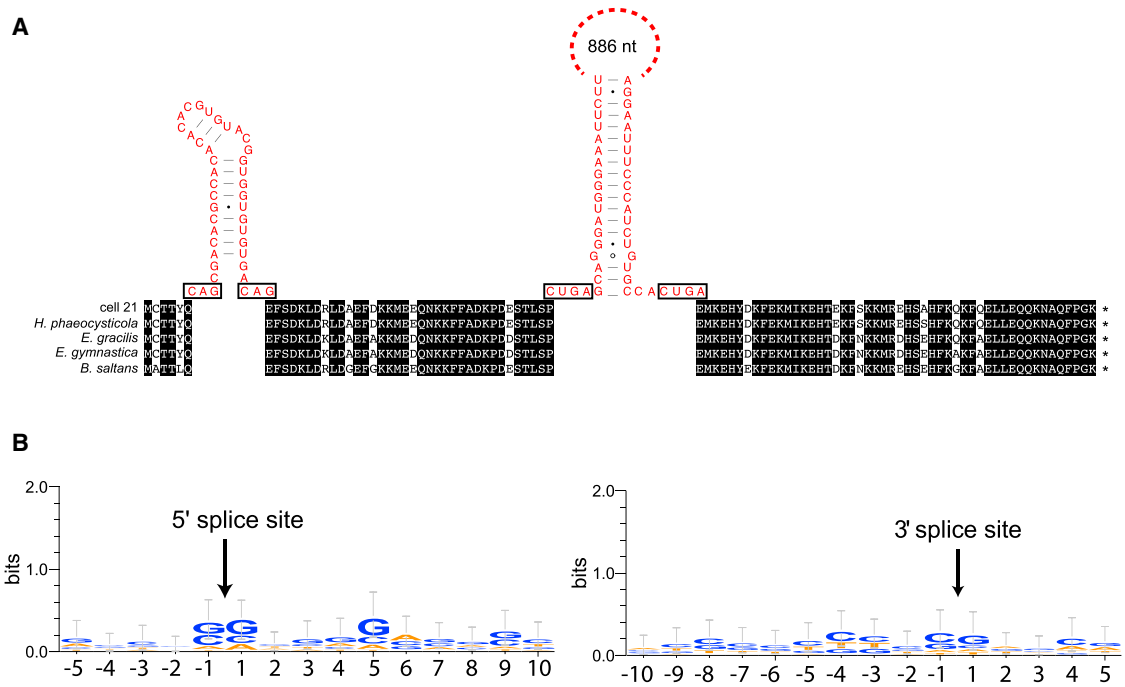


Figure 2. Marine Diplonemid Genes Contain Non-canonical Introns

(A) Multiple alignment of kinetoplastid membrane protein-11 (KMP-11) showing the position of two non-canonical diplonemid introns in cell 21 (the complete sequence of the 5′ intron is shown, but 886 nucleotides from the 3′ intron have been omitted for brevity). Extensive potential base pairing interactions between the 5′ and 3′ intron ends and short direct repeats spanning splice sites (boxed) are shown, which are both common to most observed non-canonical introns. Inferring the exact splice sites is challenging, even in this highly conserved gene, because of the direct repeats: some amino acids are (at least partially) encoded by coding sequence that could be derived from either repeat or both. In the case of the upstream intron, for example, the residues CAG must be in the mature mRNA, but three different pairs of splice sites generate the required sequence.

(B) Sequence logos of 15 residues spanning the 5′ (left) and 3′ (right) splice sites of 44 non-canonical introns (chosen due to particularly obvious capacity to form stable base pairing between 5′ and 3′ intron splice sites, to aid in their alignment). Splice sites lack canonical GT-AG boundaries but do show a slight elevation of GC.

See also Figure S2.

spliceosomal introns of other diplonemids (Figure S2) [11]. This motif is not found in the non-canonical introns but is in the 5′ intron splice sites of marine diplonemid SL RNA genes. U1 may therefore be used strictly in spliced leader addition, but canonical spliceosomal introns may also remain unobserved in marine diplonemids.

Most interestingly, the largest non-canonical intron in the SCG data—situated within an α-tubulin gene from cell 13—encodes a reverse transcriptase, a protein used for mobility of self-splicing introns in other systems, such as yeast mitochondria [16]. The non-canonical introns may therefore represent a novel and potentially still active class of mobile element that is spliced at the RNA level and, although it has existed in the euglenozoan lineage for some time, has recently spread rapidly within the marine diplonemid genomes. If so, the non-canonical introns represent an interesting and perhaps recent analogy for the origin and spread of spliceosomal introns. Regardless of their origin, they also represent a major obstacle to future genomic analyses of marine diplonemids and specifically to the generation of high-quality automated gene model predictions.

Diplonemid Ecology Inferred from SCG

Classical diplonemids are predominantly benthic and are capable of osmotrophy, phagocytosing bacteria, feeding on detritus, and (likely opportunistic) parasitism [7]. *Hemistasia* is a predator (or scavenger) of various planktonic marine eukaryotes, including haptophytes, dinoflagellates, diatoms, and copepods [17]. However, essentially nothing is known about the ecology of the marine diplonemid clade, though their increasing abundance with depth makes phototrophy unlikely [6]. The presence of reads from the SSU V9 region in unexpectedly large-size fractions has prompted suggestions that at least some are parasites of mesoplankton, perhaps in conjunction with other known parasitic groups [18].

Each of our SCG assemblies contains genomic fragments derived from prokaryotes and viruses. In some cases, obvious contamination from exogenous sources (e.g., human) was also identified. The prokaryotic signal may represent undigested prey, symbionts, co-purifying contaminants, or contaminants introduced during the sequencing process [19]. The most frequently identified bacteria were proteobacteria and cyanobacteria, typically abundant and cosmopolitan inhabitants of the open ocean [20]. On one hand, it makes sense that diplonemids would prey on these prokaryotes; however, their abundance also makes them the most likely contaminants. We also identified non-diplonemid, eukaryote-derived genes in some SCG assemblies. For instance, in cell 47, numerous contigs derived from nuclear genes of prasinophyte algae closely related

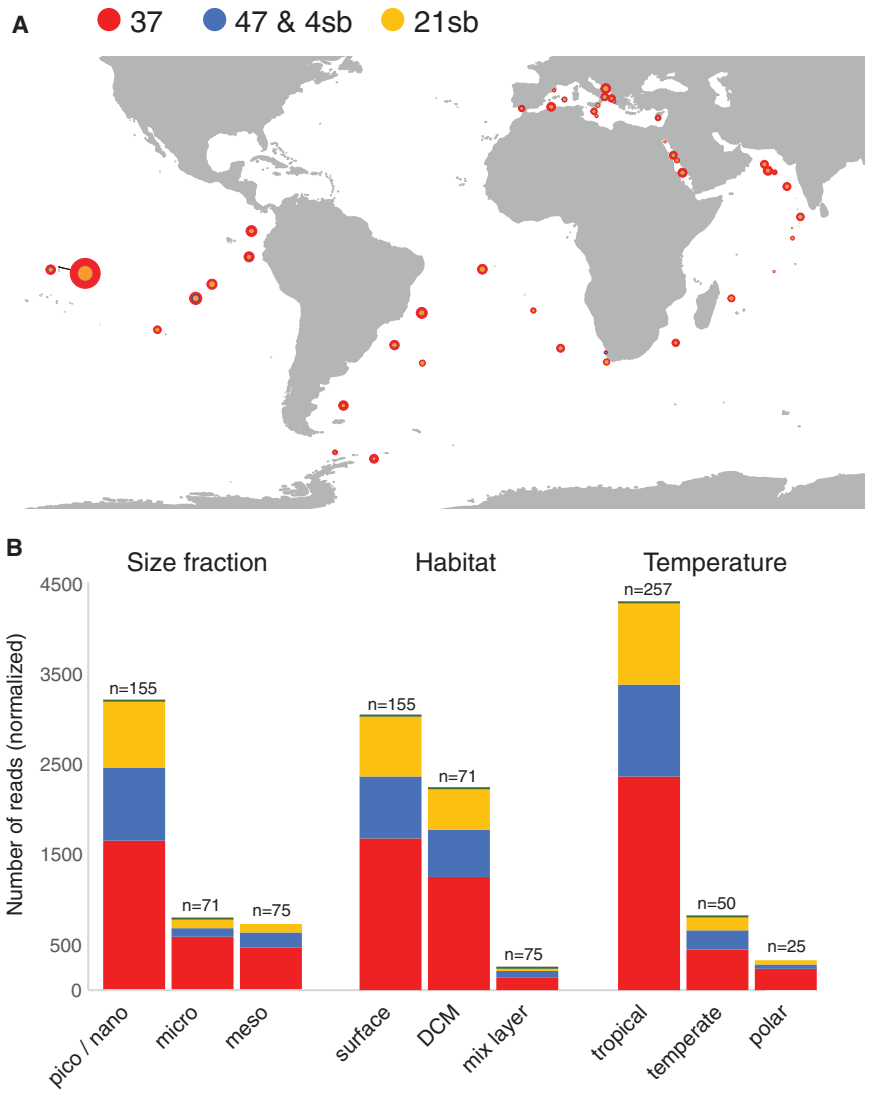


Figure 3. Geographic Distribution, Abundance, and Habitat of Marine Diplonemids Based on Tara Oceans V9 Metabarcoding Data

(A) Geographical distribution of the most-abundant phylotypes, cells 37, 47/4sb, and 21sb, based on Tara Oceans V9 metabarcoding data. Cells 47 and 4sb are pooled because they share sufficient sequence identity to be considered a single OTU, despite being isolated in different years. Dot sizes are proportional to the total number of reads in each location for that phylotype (and the phylotypes are superimposed for clarity). (B) Size fraction, habitat, and temperature distribution comparisons based on normalized Tara Oceans V9 read abundances. The color scheme for cells 37, 47/4sb, and 21sb follows (A), and less-abundant phylotypes (cells 3, 21, 27, 1sb, and 9sb) are also presented (collectively, in dark green). The number of samples per condition is presented above each chart.
See Table S4.

to *Bathycoccus* were found, along with nuclear and plastid genes derived from a haptophyte. Assemblies from cells 4sb, 37, and 21 also included likely genomic DNA from prasinophytes and haptophytes, as well as a weaker signal from photosynthetic heterokonts. Importantly, photographs of cells 47 and 4sb demonstrate the presence of one or more green bodies inside the diplonemids (Figure 1), which we interpret as partially digested prey inside a feeding vacuole. The combined photographic and genomic evidence from such "wild-caught" cells demonstrates that at least some marine diplonemids prey upon eukaryotes.

Variable Abundance of Marine Diplonemid SCG Isolates Based on V9 Data

Mapping the SSU V9 hypervariable region to Tara Oceans survey data demonstrates that the cells described here range from rare to superabundant (Figure 1). Six cells are represented by $\leq$20 Tara reads, whereas the other four are abundant (>150,000 reads) and cosmopolitan species (Figures 1 and 3A). The cells described here were most abundant in Tara pico- and nano-planktonic-size fractions (0.8–20 μm), reinforcing the idea that they are free-living predators, and they were similarly represented in surface waters and the deep chlorophyll maximum (DCM) (Figure 3B). In one case, a phylotype inferred to have relatively high abundance was isolated in both of our collection years from different depths (cells 4sb and 47; Figures 1 and 3). Most interestingly, however, is the cell 37 phylotype, which is represented by ~6,500,000 mapped reads from Tara. This accounts for 85% of reads from the entire Discoba clade, which includes not only diplonemids, kinetoplastids, and euglenoids but also heterolobosean. For perspective, the representation of this single species is greater than the total number of reads mapped to all ciliates, a whole phylum that in prior times was considered to represent abundant eukaryotic heterotrophs [1].

The Utility of SCG as a Taxonomic Framework for Interpreting Metagenomic Data

Metagenomic data have been very useful in reconstructing the roles of bacteria in complex ecosystems, such as the marine environment, but their effectiveness is entirely dependent on the quality of the reference databases on which genetic and taxonomic identifications of metagenomic assemblies are based. Eukaryotic nuclear genomes lack the breadth or depth of coverage of prokaryotes, so it has been suggested that SCG data could be an important reference tool for metagenomic analyses [21]. But with relatively few published nuclear SCG datasets from any microbial eukaryotes, it is unclear how much reference information SCG data will provide.

The marine diplonemids provide an excellent opportunity to test the performance of SCGs as references for metagenomic

data: the organisms are abundant, and we can demonstrate that they are present in well-studied systems where nuclear metagenomic data already exist. Accordingly, we mapped all reads from the Global Ocean Sampling Expedition (GOS) [4], the GOS Baltic Sea Expedition, and a survey of the Mediterranean Sea [22] to our diplonemid SCG assemblies. We found that only a tiny fraction of reads can be mapped to our nearly 1 Gb of genomic data at even modest levels of identity (e.g., at $\geq$80% or $\geq$90% identity over 100% read coverage; Table S4). The datasets investigated here are not ideal for retrieval of diplonemid sequences, as they enrich for prokaryotes, and are lacking data for greater depths, where diplonemid abundances increase [6]; however, diplonemids are not rare in surface waters [1]. We suggest that the massive underlying genetic diversity of marine diplonemids, and probably most other eukaryotic groups, coupled with the high proportion of typically fast-evolving, non-coding sequence generally found in nuclear genomes, means that, once ocean metagenomes represent the entire cross section of different protistan size classes, we should expect relatively few recognizable sequence comparisons to be present between even large metagenomic and reference SCG datasets. Individual genomes provide limited context to the vast diversity of eukaryotic metagenomic data accumulating in databases.

Conclusions

Tackling the ecological and evolutionary diversity of microbial eukaryotes is a major challenge because so little of that diversity is available in culture, and so many of the main constituents of major systems, like the marine environment, remain largely uncharacterized. The marine diplonemids typify this problem: as incredibly abundant and diverse as they are, until now they have never been visualized, and were instead known only through surveys of a single gene. By documenting the morphology and genomic characteristics of ten representatives, including one of the single-most-common heterotrophic eukaryotes in the ocean (perhaps the most common), we have a first glimpse at their morphological diversity, their co-associations, gene content, and genomic characteristics. These observations provide a guide for isolating and identifying cells for the establishment of stable cultures, which will undoubtedly yield important further insights into the evolution, ecology, and metabolism of one of the ocean's most enigmatic groups.

ACCESSION NUMBERS

The accession numbers for the raw reads and assembled scaffolds datasets reported in this paper are GenBank: SRP081436 and Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.d19j0, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.09.013.

AUTHOR CONTRIBUTIONS

N.O. isolated and photographed cells. J.F.H.S. performed MDA reactions and sequenced partial SSU genes. J.d.C. performed V9 region analysis and mapped metagenomic reads to SCG data. R.M.R.G. assembled and cleaned genomic data and performed phylogenetic and genomic analyses. J.L. provided transcriptome data from *Hemistasia*. T.A.R., A.Z.W., A.E.S., and P.J.K. conceived the study. R.M.R.G. and P.J.K. wrote the manuscript. All authors commented on and approved the manuscript.

ACKNOWLEDGMENTS

We thank the captain and crew of the R/V Western Flyer and Valeria Jiminez (co-chief scientist) as well as J. Timothy Pennington, Marguerite Blum, Christopher Wahl, Jarred Swalwell, Francisco P. Chavez, and Jason M. Smith for logistical assistance prior to and during the cruises and Akinori Yabuki, Anzhelika Butenko, Pavel Flegontov, and Aleš Horák for assistance with *Hemistasia* and discussions. This work was supported by grants from the Canadian Institutes of Health Research (MOP-42517) and from the Gordon and Betty Moore Foundation (GBMF3307) to P.J.K., T.A.R., A.Z.W., and A.E.S. Ship time was supported by a grant from the David and Lucille Packard Foundation through MBARI to A.Z.W. R.M.R.G. was supported by a fellowship from the Canadian Institutes of Health Research; J.d.C. was supported by a European Commission Marie Curie International Outgoing Fellowship grant (FP7-PEOPLE-2012-IOF – 331450 CAARL); and R.M.R.G., N.O., and J.d.C. were supported by a grant from the Tula Foundation to the Centre for Microbial Biodiversity and Evolution at UBC.

Received: June 27, 2016
Revised: August 25, 2016
Accepted: September 7, 2016
Published: November 21, 2016

REFERENCES

1. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al.; Tara Oceans Coordinators (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science *348*, 1261605.
2. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature *409*, 603–607.
3. Lara, E., Moreira, D., Vereshchaka, A., and López-García, P. (2009). Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. Environ. Microbiol. *11*, 47–55.
4. Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. *5*, e77.
5. Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015). Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science *347*, 1257594.
6. Lukeš, J., Flegontova, O., and Horák, A. (2015). Diplonemids. Curr. Biol. *25*, R702–R704.
7. Roy, J., Faktorová, D., Benada, O., Lukeš, J., and Burger, G. (2007). Description of *Rhynchopus euleeides* n. sp. (Diplonemea), a free-living marine euglenozoan. J. Eukaryot. Microbiol. *54*, 137–145.
8. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182–W185.
9. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
10. Sturm, N.R., Maslov, D.A., Grisard, E.C., and Campbell, D.A. (2001). *Diplonema* spp. possess spliced leader RNA genes similar to the Kinetoplastida. J. Eukaryot. Microbiol. *48*, 325–331.
11. Qian, Q., and Keeling, P.J. (2001). Diplonemid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer. Protist *152*, 193–201.
12. Tessier, L.H., Paulus, F., Keller, M., Vial, C., and Imbault, P. (1995). Structure and expression of Euglena gracilis nuclear rbcS genes encoding

the small subunits of the ribulose 1,5-bisphosphate carboxylase/oxygenase: a novel splicing process for unusual intervening sequences? J. Mol. Biol. *245*, 22–33.

13. Milanowski, R., Karnkowska, A., Ishikawa, T., and Zakryś, B. (2014). Distribution of conventional and nonconventional introns in tubulin (α and β) genes of euglenids. Mol. Biol. Evol. *31*, 584–593.
14. Muchhal, U.S., and Schwartzbach, S.D. (1994). Characterization of the unique intron-exon junctions of *Euglena* gene(s) encoding the polyprotein precursor to the light-harvesting chlorophyll a/b binding protein of photosystem II. Nucleic Acids Res. *22*, 5737–5744.
15. Breckenridge, D.G., Watanabe, Y., Greenwood, S.J., Gray, M.W., and Schnare, M.N. (1999). U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. Proc. Natl. Acad. Sci. USA *96*, 852–856.
16. Kennell, J.C., Moran, J.V., Perlman, P.S., Butow, R.A., and Lambowitz, A.M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. Cell *73*, 133–146.
17. Yabuki, A., and Tame, A. (2015). Phylogeny and reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrächter & Schnepf, 1996. J. Eukaryot. Microbiol. *62*, 426–429.
18. Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al.; Tara Oceans Coordinators (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. Science *348*, 1262073.
19. Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S., and Bhattacharya, D. (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science *332*, 714–717.
20. Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature *420*, 806–810.
21. del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014). The others: our biased perspective of eukaryotic genomes. Trends Ecol. Evol. *29*, 252–259.
22. López-Pérez, M., Kimes, N.E., Haro-Moreno, J.M., and Rodriguez-Valera, F. (2016). Not all particles are equal: the selective enrichment of particle-associated bacteria from the Mediterranean Sea. Front. Microbiol. *7*, 996.