

---

4-1-2018

## Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data

Jean David Grattepanche  
*Smith College*

Laura M. Walker  
*Smith College*

Brittany M. Ott  
*University of Maryland, College Park (UMD)*

Daniela L. Paim Pinto  
*University of Rhode Island*

Charles F. Delwiche  
*University of Maryland, College Park (UMD)*

*See next page for additional authors*

Follow this and additional works at: [https://scholarworks.smith.edu/bio\\_facpubs](https://scholarworks.smith.edu/bio_facpubs)



Part of the [Biology Commons](#)

---

### Recommended Citation

Grattepanche, Jean David; Walker, Laura M.; Ott, Brittany M.; Paim Pinto, Daniela L.; Delwiche, Charles F.; Lane, Christopher E.; and Katz, Laura A., "Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data" (2018). Biological Sciences: Faculty Publications, Smith College, Northampton, MA.  
[https://scholarworks.smith.edu/bio\\_facpubs/91](https://scholarworks.smith.edu/bio_facpubs/91)

This Article has been accepted for inclusion in Biological Sciences: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact [scholarworks@smith.edu](mailto:scholarworks@smith.edu)

---

**Authors**

Jean David Grattepanche, Laura M. Walker, Brittany M. Ott, Daniela L. Paim Pinto, Charles F. Delwiche, Christopher E. Lane, and Laura A. Katz

# Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data

Jean-David Grattepanche, Laura M. Walker, Brittany M. Ott, Daniela L. Paim Pinto, Charles F. Delwiche, Christopher E. Lane, and Laura A. Katz\*

Despite their diversity and ecological importance, many areas of the SAR—Stramenopila, Alveolata, and Rhizaria—clade are poorly understood as the majority (90%) of SAR species lack molecular data and only 5% of species are from well-sampled families. Here, we review and summarize the state of knowledge about the three major clades of SAR, describing the diversity within each clade and identifying synapomorphies when possible. We also assess the “dark area” of SAR: the morphologically described species that are missing molecular data. The majority of molecular data for SAR lineages are characterized from marine samples and vertebrate hosts, highlighting the need for additional research effort in areas such as freshwater and terrestrial habitats and “non-vertebrate” hosts. We also describe the paucity of data on the biogeography of SAR species, and point to opportunities to illuminate diversity in this major eukaryotic clade. See also the video abstract here: [https://youtu.be/\\_VUXqaX19Rw](https://youtu.be/_VUXqaX19Rw).

blooms], human health, microbiomes).<sup>[3–14]</sup> The bulk of eukaryotes are microbial; plants, animals, and fungi represent only three of an estimate of 75–100 major clades. Within microbial eukaryotes, the SAR clade—Stramenopila, Alveolata, Rhizaria—contains a tremendous diversity of lineages, with at least 60 944 named species to date. Numerous analyses based on microscopy and more recently molecular data show a substantial contribution of the SAR lineage to microbial eukaryotic communities in diverse environments.<sup>[15–24]</sup> The SAR lineage represents a large part of the eukaryotic diversity and is estimated to be as large as 50% of eukaryote diversity.<sup>[25]</sup> As we enter a period of accelerated “omics” data collection, it is essential to understand the state of knowledge about this diverse clade.

## 1. Introduction

Microbes were the first and only inhabitants of Earth for almost two thirds of our planet’s history.<sup>[1,2]</sup> Microbes are present everywhere, are important players in ecosystem processes (e.g., in food webs and biogeochemical cycles), and have substantial impacts on our global socio-economy (e.g., fisheries, agriculture, tourism [toxic

The eukaryotic clade “SAR” unites Stramenopila, Alveolata, and Rhizaria,<sup>[26]</sup> and contains an immense diversity of lineages that represent different morphologies (e.g., amoebae, ciliates, flagellates), live almost everywhere (e.g., marine, freshwater, soil, symbionts), and include many important parasites of animals (e.g., *Plasmodium*, the causative agent of malaria) and plants (e.g., Peronosporomycetes, or water molds, including the species responsible for the Irish Potato famine). The monophyly of SAR emerged from multi-gene phylogenies<sup>[26]</sup> and has been repeatedly recovered,<sup>[14,27,28]</sup> although the specific relationships among the three lineages varies across analyses. Even after molecular analyses indicated a recent common ancestor for these lineages, no morphological or ultrastructural synapomorphy has been identified. Fundamentally, the uniting of these clades rejects the controversial but influential “Chromalveolate hypothesis,”<sup>[29]</sup> which argued for a single primary acquisition of red algal plastids in the last common ancestor of “chromalveolates” (Stramenopila, Alveolata, Haptophyta, and Cryptomonads). Instead, the SAR clade includes Rhizaria, and excludes Haptophyta and Cryptomonads. Therefore, uniting the three morphologically-diverse clades of stramenopiles, alveolates, and rhizaria as a monophyletic group has broad implications for our understanding of eukaryotic evolution and the evolution of photosynthesis.

The variety of organisms within SAR is indeed tremendous. For example, SAR includes important photosynthetic lineages such as diatoms and kelp (Stramenopila), pathogenic parasites

Dr. J.-D. Grattepanche, Dr. L. M. Walker, Prof. L. A. Katz  
 Department of Biological Sciences  
 Smith College,  
 Northampton MA 01063, USA  
 E-mail: lkatz@smith.edu

Dr. B. M. Ott, Prof. C. F. Delwiche  
 Department of Cell Biology and Molecular Genetics  
 University of Maryland  
 College Park MD 20742, USA

Dr. Daniela L. Paim Pinto, Prof. C. E. Lane  
 Department of Biological Sciences  
 University of Rhode Island  
 Kingston RI 02881, USA

© 2018 The Authors. *BioEssays* Published by WILEY Periodicals, Inc. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: 10.1002/bies.201700198

such *Plasmodium* and *Toxoplasma* (Alveolata), as well as Foraminifera (Rhizaria) that build complex “shells” (tests) ranging from microns to centimeters that are fossils and serve as bioindicators. These exemplar taxa represent organisms that demonstrate novel features and/or are economically important (i.e., foraminiferans are used to identify oil deposits (e.g., Ref. [30]), and therefore are among some of the best studied groups within SAR. However, despite the data that exist for a small number of lineages, the large majority of SAR lineages remain substantially understudied.<sup>[25]</sup> To synthesize data on this major eukaryotic clade, we provide an update on the current state of the knowledge regarding SAR and highlight some of the implications that the monophyly of SAR has on our understanding of the evolutionary history of eukaryotes. We also provide data on the proportion of SAR that remains underexplored, the ecology of SAR lineages and the biogeography of exemplar species. Finally, we provide insight for future investigation as the community works to fill in knowledge from diverse and understudied SAR lineages.

## 2. What is Known About the Three Major Lineages Within SAR?

Each lineage of the SAR clade has been reviewed elsewhere (e.g., Refs. [31–34]) and are a focus in numerous books (e.g., Refs. [21,35–37]). Here we summarize what is known about the SAR clade, focusing on all three lineages (Stramenopila, Alveolata, and Rhizaria; e.g., Refs. [26,38–40]). We highlight synapomorphies when present, indicate understudied taxa and give a brief overview of the incredible diversity of each clade.

### 2.1. Stramenopila is Species Rich, Particularly for Phototrophs

The monophyly of Stramenopila (a name proposed by Patterson<sup>[41]</sup> and concordant with the clade Heterokonta<sup>[42]</sup>) emerged from both morphological work<sup>[42,43]</sup> and molecular studies, initially of the SSU-rDNA gene.<sup>[44,45]</sup> The members of the group are characterized by the presence of a long, anteriorly-directed flagellum that has tiny tripartite bristles called mastigonemes (also called stramenopili) that help propel the cell forward (**Figure 1**). Most species also present a shorter, smooth posterior flagellum (hence the alternative name “Heterokonts”) used in a whip-like motion to steer the swimming direction.<sup>[46–49]</sup> While predominantly photosynthetic, numerous lineages lack plastids including the early-diverging opalinids, oomycetes, and labyrinthulids (see below,<sup>[50]</sup>).

Roughly 30 000 species are recognized in the Open Tree Taxonomy<sup>[51]</sup> within the Stramenopila, but as many as 1 000 000 species have been estimated to exist,<sup>[52]</sup> and most of the biodiversity is believed as yet to be unidentified. Recent reports have consistently unveiled novel clades using a variety of sequencing and taxonomic approaches and largely focusing on marine habitats.<sup>[53–60]</sup> Despite their importance as primary producers (fixing a considerable amount of CO<sub>2</sub>) and in other biogeochemical cycles (e.g., nitrogen and sulfur cycling), there remains a lack of molecular information for most stramenopiles

lineages, resulting in poor support for phylogenetic relationships along the backbone of this massive clade.<sup>[61,62]</sup>

The Stramenopila comprise a monophyletic cluster of autotrophic organisms (Ochrophyta) and several heterotrophic lineages with a wide variety of life history strategies.<sup>[50,54,55,63]</sup> These heterotrophic clades include several unicellular members (e.g., bicosoecids, “marine stramenopiles” [MAST], placidids), parasitic species (e.g., *Blastocystis*, Peronosporomycetes, labyrinthulids), and endocommensals (e.g., Opalinida). The diversity of heterotrophic stramenopiles is poorly understood, and environmental sequencing efforts have revealed several new lineages over the past 15 years (e.g., Refs. [56,64–67]), many of which have not been directly observed.

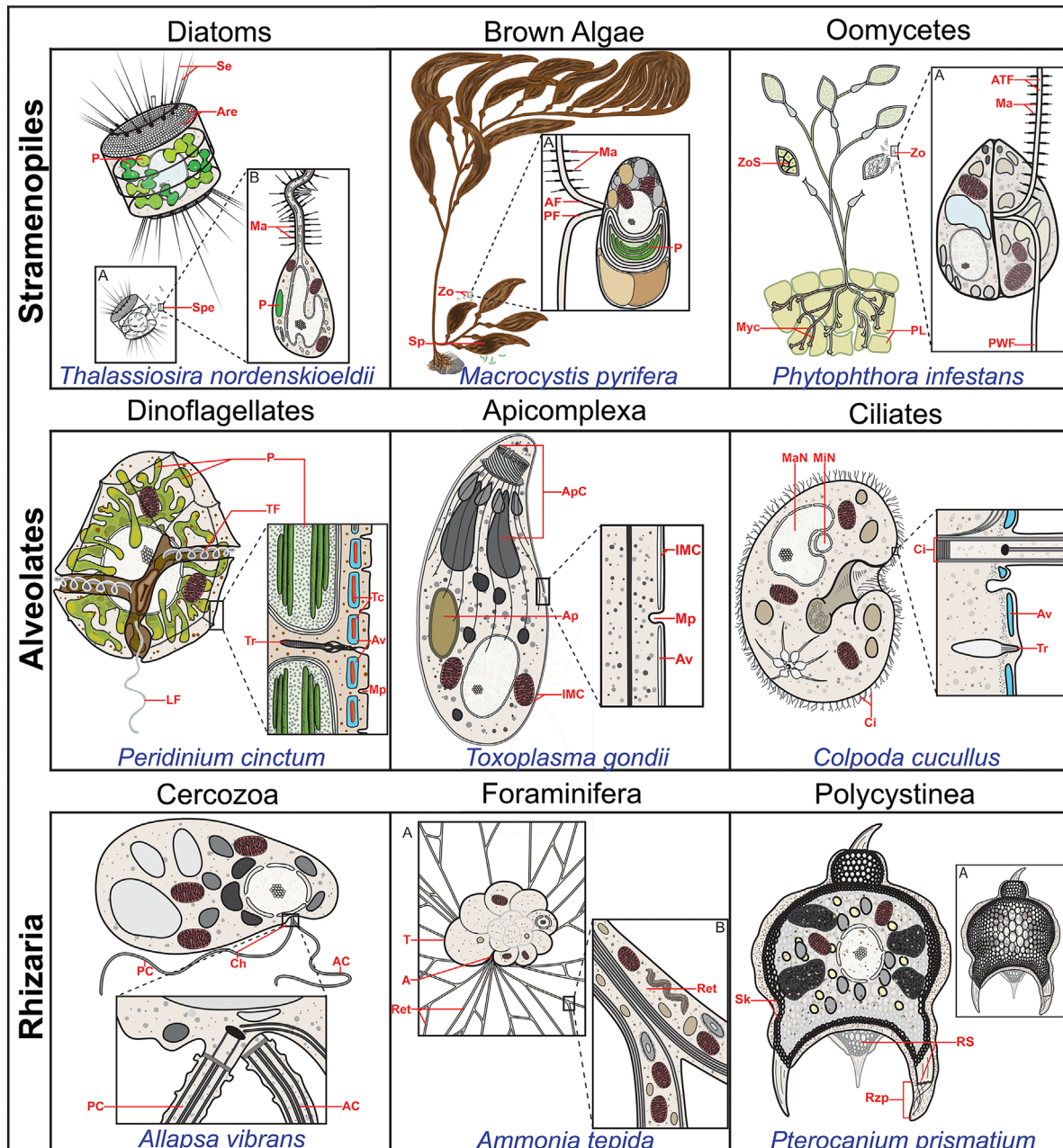
The autotrophic stramenopiles are far more familiar and include giant multicellular algae (e.g., kelp, phaeophytes), lineages with a mix of single-celled and multicellular species (e.g., xanthophytes) and those dominated by single-celled algae (e.g., diatoms, ochromonads). The kelp forests (**Figure 1**) on the west coast of North America are probably the most charismatic stramenopiles, but diatoms, with their intricate glass cases, are held in similar regard.<sup>[50,52,68]</sup> The outer frustules of the diatoms and scales of the chrysophytes fossilize well and have made them critical indicators of historical climate, both recent,<sup>[69]</sup> and ancient.<sup>[70]</sup>

### 2.2. Alveolata is a Morphologically and Ecologically Diverse Clade

The Alveolata<sup>[71]</sup> accommodate the grouping of ciliates with dinoflagellates and Apicomplexa. Similar to stramenopiles, this classification was initially based on SSU-rDNA phylogeny,<sup>[71,72]</sup> has been consistently recovered in numerous phylogenetic analyses, and hence is broadly accepted as a monophyletic lineage.<sup>[32,73,74]</sup> Cavalier-Smith<sup>[71]</sup> proposed that the key morphological feature uniting the organisms was the presence of cortical alveoli (membrane-bound sacs underlying the cell membrane, **Figure 1**), and on that basis suggested the name Alveolata. Currently this clade contains the three well-known lineages, Dinoflagellata, Apicomplexa, and Ciliophora, as well as numerous smaller groups with uncertain placements (e.g., Perkinsidae, Chromerida, and Colpodellida) comprising a total of ≈17 000 named species in the Open Tree Taxonomy.<sup>[51]</sup>

Alveolates are mostly unicellular organisms, although some species form filaments or chains, and still others (particularly ciliates like *Stentor* and *Blepharisma*) have remarkably large and complex cells that can reach millimeters in length. A diverse array of trophic modes is found among the alveolates including parasitic (e.g., apicomplexans), heterotrophic (e.g., ciliates, some dinoflagellates), autotrophic (e.g., some dinoflagellates), and mixotrophic (with plastids that may be permanent or kleptoplastidic, e.g., some dinoflagellates and ciliates) lineages.

Ciliates are one of the earliest microbes to be described, having appeared in the published literature as early as 1867.<sup>[75,76]</sup> Ciliates are characterized by the presence of cilia and dimorphic nuclei,<sup>[77]</sup> and have both a somatic macronucleus (highly processed, amplified, and transcriptionally active) and a germline micronucleus (transcriptionally silent) present in every cell (**Figure 1**). The tremendous morphological diversity of ciliates



**Figure 1.** Illustration of representative lineages within SAR (Stramenopila, Alveolata, and Rhizaria), including synapomorphies where they exist: flagella with tripartite hairs in Stramenopila, alveolar sacs in Alveolata. In Diatoms, A) shows a male cell releasing sperm, the structure of which is shown in detail B). In Brown Algae, A) shows a zoospore, which has been released from the sporophytes. In Oomycetes (or Peronosporomycetes), A) shows a zoospore, which has been released from a zoosporangium. In Alveolata, the inset shows a detailed drawing of the cell membrane/wall. In Cercozoa, the inset shows the ciliary hub (Ch). In Foraminifera, A) shows a dorsal view of the cell, with reticulopodia (detailed in B)) emerging from the aperture. In Polycystinea, we see an event of wound-healing performed by the rhizopods. Abbreviations: A-Aperture; AC-Anterior cilium; AF-Anterior flagellum; AP-Apicoplast; ATF-Anterior Tinsel Flagellum; AV-Alveoli/us; Are-Areolae; IMC-Inner Membrane Complex; LF-Longitudinal flagellum; Ma-Mastigonemes; MaN-Macronucleus; MiN-Micronucleus; Mp-Micropore; Myc-Mycellum; P-Plastid; PC-Posterior cilium; PF-Posterior flagellum; PL-Host plant cell; PWF-Posterior whiplash flagellum; Ret-Reticulopodia; RS-Radial spines; Rzp-Rhizopod; Se-Setae; Sk-Skeleton; Sp-Sporophyll; Spe-Sperm; T-Test; Tc-Thecal plate; TF-Transverse flagellum; Tr-Trichocyst; Zo-Zoospore; ZoS-Zoosporangium. For greater detailed, full-sized images, please see Supporting Information.

has led to the identification of roughly  $\approx 8000$  species that are currently placed among 11 classes.<sup>[31]</sup> Beyond their morphological diversity, ciliates such as *Tetrahymena thermophila* are often used as model organisms for both cell and molecular biology, and have allowed numerous influential discoveries such as the discovery of telomeres and telomerases,<sup>[78]</sup> self-splicing introns,<sup>[79]</sup> and many epigenetic processes.<sup>[80–83]</sup>

Approximately half of the dinoflagellates are photosynthetic and the remainder includes both predatory and parasitic species. Many species are mixotrophic and able to change their nutritional mode depending upon environmental circumstances. Dinoflagellates can either be armored (i.e., thecate, Figure 1) or naked,<sup>[84,85]</sup> and are abundant in both marine and freshwater environments.<sup>[51]</sup> Because they are the causative agents of many “red tides” and are sometimes toxin-producing, dinoflagellates have considerable economic relevance in coastal communities.<sup>[86]</sup> Moreover, dinoflagellates in partnership with Cnidaria (a clade of animals including the great star coral *Montastraea cavernosa*) create the coral reefs that support some of the most diverse and productive marine habitats.<sup>[87,88]</sup> Most dinoflagellates also have unusual genomes that are organized into permanently condensed chromosomes that lack nucleosomes.<sup>[89,90]</sup>

The Apicomplexa, named for the distinctive and elaborate apical complex that facilitates host penetration, were long thought to be exclusively animal parasites. Surprisingly, many species harbor non-photosynthetic plastids<sup>[91,92]</sup> termed apicoplasts (Figure 1). The presence of apicoplasts, combined with their sister relationship to the recently discovered, photosynthetic Chromerids,<sup>[93]</sup> suggests a photosynthetic ancestor for this clade.<sup>[93,94]</sup> The causative agents of malaria, found in the genus *Plasmodium*, are among the most important apicomplexan parasites but other devastating apicomplexan diseases include toxoplasmosis, babesiosis, and coccidiosis. Recently, free-living apicomplexans have been identified in the oceans<sup>[15]</sup> and in the soil of Neotropical forests,<sup>[16]</sup> demonstrating that only a fraction of their diversity is characterized.

### 2.3. The Recently Characterized Rhizaria Still Lacks a Morphological Synapomorphy

Rhizaria, a recently proposed major clade of eukaryotes, contains a tremendous diversity of lineages, many of which are poorly understood.<sup>[95]</sup> For example, numerous lineages (e.g., “radiolarians,” Foraminifera, euglyphids) remain under-sampled with molecular techniques (see below) and no morphological synapomorphy has yet been identified for the group as a whole.<sup>[95,96]</sup> The majority of rhizarian species are heterotrophic, free-living, and are difficult to isolate and culture. Instead, more attention has been paid to the few parasitic taxa (albeit, no human pathogens) such as *Plasmodiophora brassica*, which causes club root in cruciferous vegetables. The few photosynthetic lineages in Rhizaria show clear sign of secondary endosymbiosis (i.e., of a green algae symbiont in chlorarachniophytes<sup>[97,98]</sup>) or recent independent acquisition of a cyanobacterial symbiont (i.e., in *Paulinella chromatophora*<sup>[99]</sup>). Rhizaria also contains several amoeboid lineages (e.g., Foraminifera and “euglyphids”) that have filose (fine) pseudopods varying from simple to branching and anastomosing (Figure 1).

The monophyly of Rhizaria emerged entirely from molecular studies, and is supported by numerous multi-gene phylogenies.<sup>[26,100–102]</sup> The first evidence unifying members of Rhizaria came from analyses of SSU-rDNA that united the photosynthetic chlorarachniophytes with heterotrophic testate “euglyphid” amoebae.<sup>[103]</sup> One of the most striking morphologies found in Rhizaria are the intricate tests and complex pseudopodial networks found in “radiolarians” (i.e., the non-monophyletic grouping of Polycystinea and Acantharia), “euglyphids” (another non-monophyletic group), and Foraminifera, respectively (Figure 1). The “radiolarians” have been recognized largely for their beautiful tests, first popularized by drawings of Ernst Haeckel,<sup>[104]</sup> while foraminiferans are best known by their rich fossil record and utility in paleoecological applications. Less attention, however, has been paid to the role of these organisms in global biogeochemical cycling and recent studies indicate that large marine Rhizaria represent a significant proportion of the global planktonic community<sup>[105]</sup> and are involved in a substantial proportion of the oceans biotic carbon stock.<sup>[106]</sup>

Rhizaria are also important members of freshwater and soil microhabitats. Particularly of note is the abundance of Cercozoa in some soils where two lineages, glissomonads, and cercozoans, represent an estimated 30–60% of the total protist community (Figure 1; <sup>[107–110]</sup>). Also intriguing is a recent finding that the large, testate amoebae of Euglyphidae, usually considered aquatic, are also abundant soil inhabitants wherein they have been documented to constitute approximately 19% of the total soil SSU-rDNA lineages.<sup>[108]</sup> Given their abundance and comparatively large size (avg.  $\approx 75 \mu\text{m}$ ), these data indicate a previously underappreciated role of Rhizaria in the soil microhabitat.<sup>[108,111]</sup>

There are currently  $\approx 12\,000$  named Rhizaria species in the Open Tree Taxonomy,<sup>[51]</sup> but this number is largely underestimated given the large number of environmental rhizarian sequences readily detected from environmental samples. Moreover, a large proportion of Rhizaria are undoubtedly missed in environmental surveys given their unusual rDNAs. The SSU-rDNA in Foraminifera, for example, contain numerous insertions and substitutions<sup>[112]</sup> making them resistant to PCR with universal primers.

## 3. Knowledge of the SAR Clade Illuminates the Evolution of Photosynthesis in Eukaryotes

The monophyly of SAR changed our understanding of the evolutionary history of plastid acquisition across the eukaryotic tree of life. One popular hypothesis about the secondary spread of plastids across eukaryotes, based largely on the assumption that transfer of plastids must be rare, was that two independent lineages had acquired green algae as plastids (i.e., chlorarachniophytes and euglenids), and that a single event involving a red alga occurred in the ancestor of the “chromalveolates,” which included lineages (i.e., Stramenopila, Alveolata, Haptophyta, and Cryptomonads) with similarly pigmented plastids thought to be derived from red algae.<sup>[113]</sup> The “Chromalveolate hypothesis” was never fully consistent; in particular, the presence of many early branching, non-photosynthetic lineages (e.g., in the stramenopiles and dinoflagellates) are difficult to reconcile with

the hypothesis.<sup>[114]</sup> As multiple publications began to identify SAR as a clade (grouping Stramenopila, Alveolata, and Rhizaria, and excluding Haptophyta and Cryptophyta), the idea that plastids were present in the common ancestor became increasingly implausible.<sup>[26,101,102,115]</sup>

The distribution of plastids among SAR requires multiple gains and losses. Several hypotheses have emerged that invoke primary, secondary, and even tertiary transfers of the red algal secondary plastid,<sup>[116–121]</sup> but few have generated unified support. It seems likely that the common ancestor of dinoflagellates and apicomplexan was equipped with a plastid, despite the fact that many dinoflagellates are not photosynthetic. Most photosynthetic dinoflagellates have a red algal-derived plastid with distinctive pigmentation (peridinin) and fix carbon with an unusual enzyme (form II RuBisCO), which has otherwise only been identified in bacteria living in low-oxygen environments. Remarkably, the recently discovered Chromerida, proposed as a sister group to Apicomplexa, are photosynthetic and fix carbon with the same enzyme as photosynthetic dinoflagellates.<sup>[122]</sup> This strongly suggests that the plastid of parasitic apicomplexan is homologous to that of the photosynthetic dinoflagellates, which in turn implies that the common ancestor of Apicomplexa and Dinoflagellata was photosynthetic. Under this model, the deep branching, non-photosynthetic (predatory or parasitic) dinoflagellate lineages must have lost their plastids. There is no clear evidence of an ancestral plastid in any ciliate, suggesting either that plastids were acquired in the

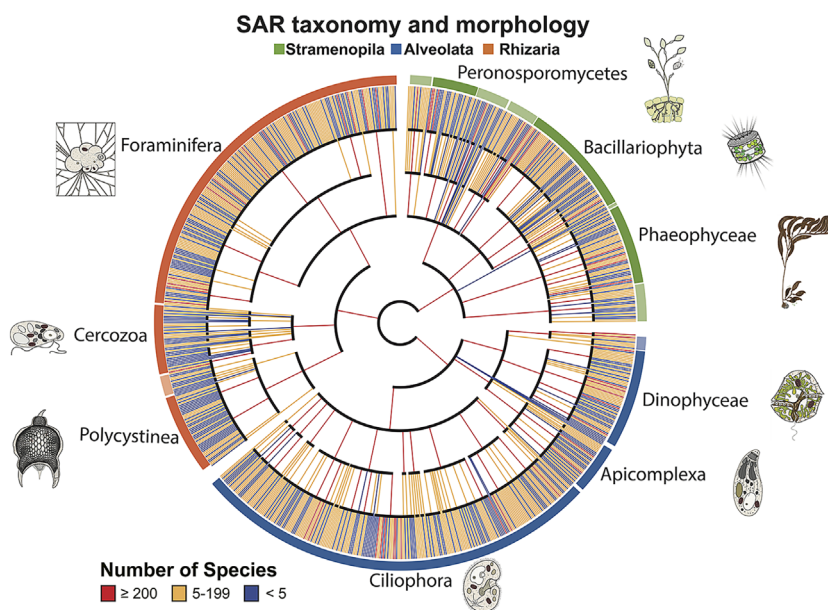
common ancestor of dinoflagellates and Apicomplexa, or that ciliates lost their plastid (and associated genes) early in their evolution.

In Stramenopila, photosynthetic lineages are highly diverse, although in most recent phylogenetic analyses they form a monophyletic group sister to the non-photosynthetic Peronosporomycetes. Sister to that clade is a large, non-monophyletic cluster of non-photosynthetic organisms including Labyrinthulea, and Opalozoa.<sup>[53,56,62]</sup> There is little evidence of ancestral plastids in any of the non-photosynthetic stramenopiles, which, taken together with the branching order, implies that the common ancestor was not photosynthetic. Although the pigmentation of stramenopile plastids is superficially similar to that of dinoflagellates, they are structurally quite different, suggesting independent acquisition of secondary plastids from red algae. There is no compelling evidence of plastids in the last common ancestor of Rhizaria.

#### 4. Analyses of the SAR Clade Reveals a Dark Area Between Taxonomy and Molecular Data

Although the majority of information we have presented so far relies on taxonomy (i.e., named species, mostly based on morphological analyses), we further assessed the state of knowledge of the SAR lineages by analyzing available molecular data on GenBank. Using the taxonomy from Open Tree of Life (Open Tree Taxonomy;<sup>[51]</sup>), we identified 60 944 species (i.e., lineages with binomial names) within SAR including 31 813 Stramenopila, 17 229 Alveolata, and 11 902 Rhizaria (Figure 2; Table 1; Table S1, Supporting Information). The taxonomy reveals a bias related to the “age” of the first description. For example, the long-recognized ciliate clade is composed by many classes and orders while the more recently described Cercozoa clade has only a few ranks (Figure 2). We also note that approximately 11% of these species are labeled as extinct in the Open Tree Taxonomy, and therefore will not have molecular data (Figure 3).

To evaluate availability of molecular data, we chose the SSU-rDNA (small subunit ribosomal RNA) gene because it is the “gold standard” for many studies of microbial diversity; it is the most commonly sequenced gene for the highest number and largest diversity of taxa (e.g., Refs. <sup>[123–129]</sup>). GenBank contains more than 30 000 unique SSU-rDNA sequences for SAR lineages corresponding to 21 041 records for named species (taxa with a binomial name), and the remainder having been submitted as “environmental” or “uncultured” (see Supporting Information methods for more details). These 21 041 records range between 500 and 3500 bp in length and represent 4011 SAR species (see Supporting Information methods). We also observed a



**Figure 2.** SAR taxonomic tree. Analyses of SAR taxonomy show a tremendous number of species (60 944 binomial names) and an unequal taxonomy for each lineage (e.g., the long-recognized Ciliophora clade has many classes and orders while the more recently described Cercozoa clade is composed of few ranks). The taxonomy is built from the Open Tree Taxonomy version 3.0 (<https://tree.opentreeoflife.org/about/taxonomy-version/ott3.0> access on 2 November 2016). The circular tree is limited to five ranks (from inside to outside ring of the tree: “infrakingdom,” phylum, class, order, and family). Major clades are indicated on the edges of the tree, warmer colored branches indicate greater numbers of descendant species. More details can be found in Supporting Information.

**Table 1.** Summary of data on major SAR lineages from the Open Tree Taxonomic and GenBank databases used for the Figures 1–3.

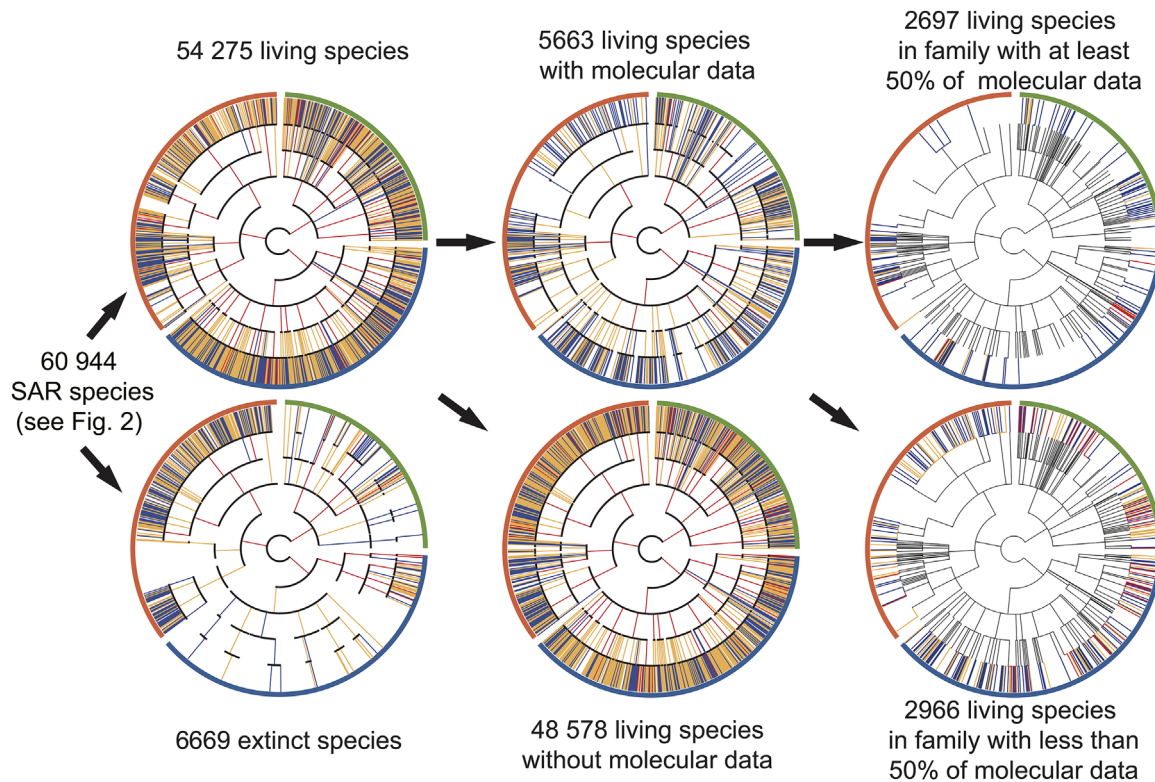
SAR Clade	Clade	Common name	Species in OTT	Species in GB	Records in GB
Stramenopila	Bacillariophyta	Diatoms	23 542	806 (3%)	2010
	Phaeophyceae	Brown algae	2875	120 (4%)	221
	Peronosporomycetes	Water molds	2445	183 (7%)	958
	Others		2951	335	1708
Alveolata	Ciliophora	Ciliates	9648	1036 (11%)	2276
	Apicomplexa		1132	524 (46%)	6639
	Dinoflagellata	Dinoflagellates	5158	424 (8%)	1354
	Others		1291	17	86
Rhizaria	Foraminifera		10 481	257 (2%)	5128
	Cercozoa		691	183 (27%)	399
	Polycystinea + Acantharia	Radiolioria	724	97 (13%)	215
	Others		6	29	47

Full data in Table S1, Supporting Information.

bias toward alveolates in terms of the number of sequences (or records) in GenBank. Rhizaria and Stramenopila each represent around 25% of the SAR records while the remaining records ( $\approx 50\%$ ) are Alveolata (10 355 records), in particular apicomplexan records (Table 1).

#### 4.1. What is the Darkness Within the SAR Clade?

Using the Open Tree Taxonomy and GenBank databases, we distinguished “dark areas” (i.e., species lacking molecular data) and “well lit” areas (groups well studied using molecular tools;

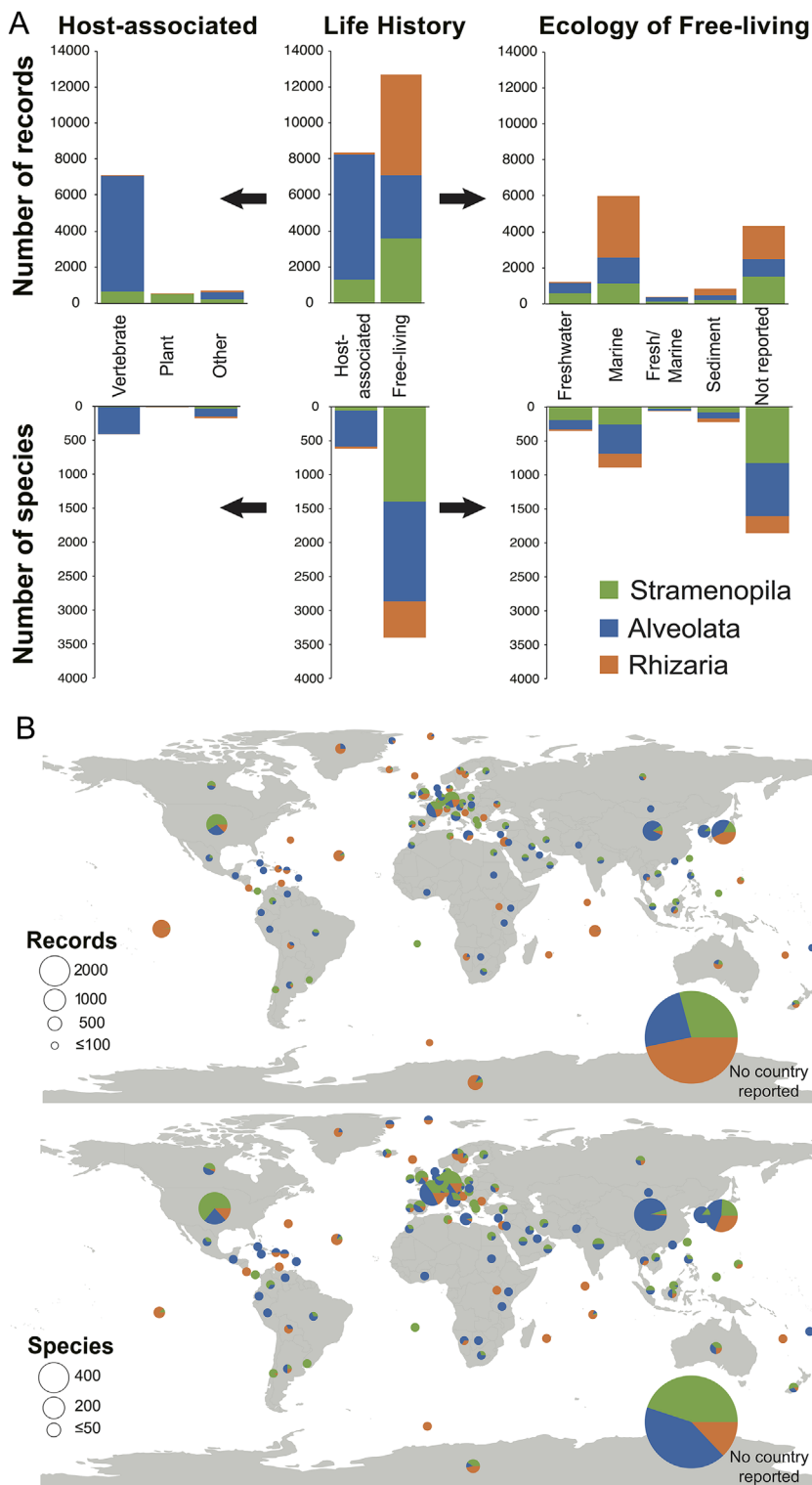


**Figure 3.** Dark area in the SAR clade. Analyses of SAR taxonomy show a tremendous number of species (60 944 binomial names) of which 6669 are extinct, only 5663 have any molecular data, and less than 5% are “well-known” (i.e., species from a family with at least 50% of the species having molecular data). The taxonomy is built from the Open Tree Taxonomy version 3.0 (<https://tree.opentreeoflife.org/about/taxonomy-version/ott3.0> access on 2 November 2016). The circular tree is limited to five ranks (from inside to outside ring of the tree: “infrakingdom,” phylum, class, order, and family). Major clades are indicated on the edges of the tree, warmer colored branches indicate greater numbers of descendant species.



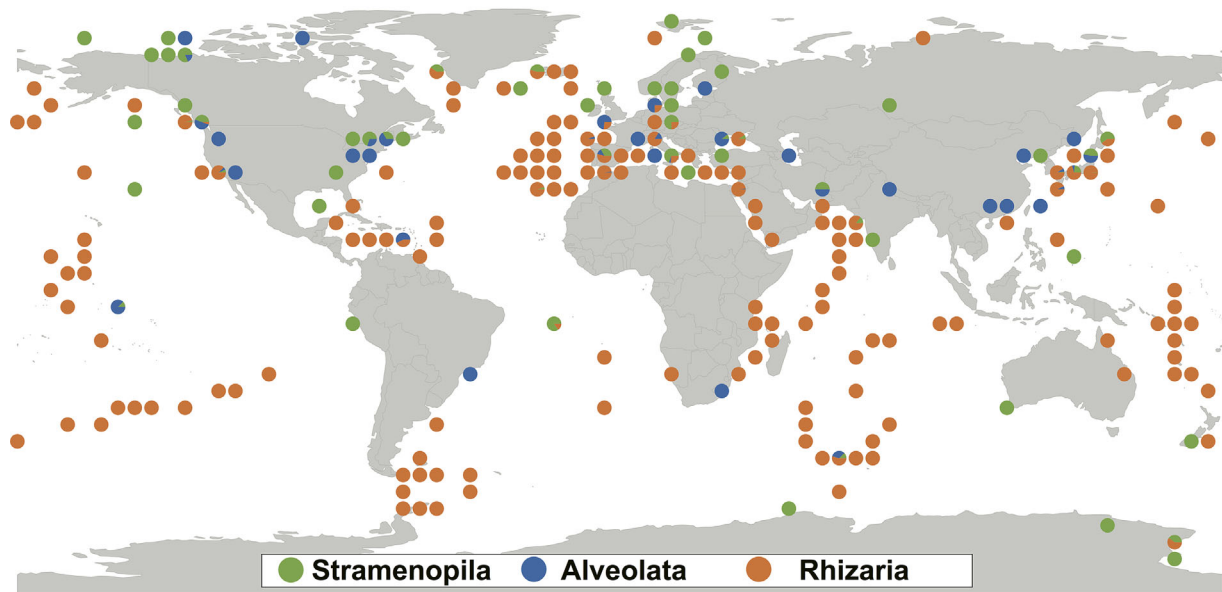
Figure 3; see Supporting Information for methods) of SAR. Almost 90% of named living species lack molecular data (SSU-rDNA or other markers), indicating that the bulk of diversity is described only by morphology (48 578 living species without molecular data; Figure 3). We also evaluate the groups most thoroughly studied, which we identified as families in which more than 50% of the species have been characterized using molecular data. Using these criteria, we find that the best studied groups include: parasites such as the Apicomplexa (Alveolata) and Mikrocystiidae (Rhizaria), macroscopic lineages such as the brown algae, (Phaeophyceae, Stramenopila), and golden algae (Chrysophyceae, Stramenopila; Figure 3). A large number of cercozo (Rhizaria) have also been well characterized by sequencing, which is likely due to the fact that this relatively recently characterized clade<sup>[130]</sup> has largely been studied using molecular data (e.g., Ref. <sup>[131]</sup>), and lacks a long history of taxonomic study (Figure 3). Clades that have substantially fewer species assessed using molecular tools include Foraminifera (Rhizaria), “radiolarians” (i.e., Polycystinea and Acantharia [Rhizaria]), many ciliate lineages (Alveolata), and the species rich diatoms (i.e., Bacillariophyta, Stramenopila; Figure 3).

Of the 5663 Open Tree Taxonomy species with molecular data, 70% have been characterized by their SSU-rDNA in GenBank. The remaining Open Tree Taxonomy species have been characterized by their LSU-rDNA (large subunit ribosomal RNA), plastid or mitochondrial sequences, or other genes. This includes less than 10% of Rhizaria species (566 species), around 25% of the Stramenopila (1444 species), more than 35% (2001 species) of Alveolata species (Table 1). The distribution of sequences within each clade matches the pattern described above for total molecular data, wherein macroscopic taxa, as well as medically and economically relevant taxa, have been most heavily investigated by molecular data (Figure 3). Confounding this analysis is the well-known problem of plurality of species concepts, particularly between morphology and molecular phylogeny (reviewed in<sup>[132]</sup>). We observe that: (1) fewer than 8% of the SAR species have been assessed by molecular data and (2) there is a bias toward the alveolates in terms of species and record number, particularly toward the apicomplexans. In the following sections, we assess the various biases and the ecology of the SAR clade, in particular life history, ecology, and biogeography associated with the GenBank records.

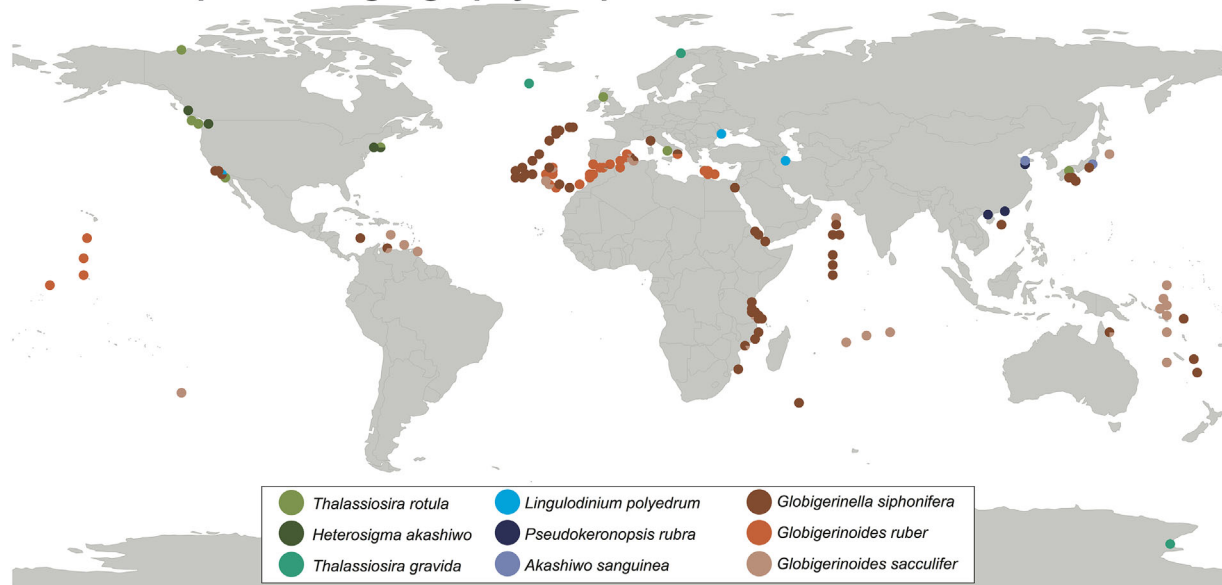


**Figure 4.** Distributions of SAR species from metadata associated with GenBank records, depicted by (A) life history and ecology categories and by (B) country. This analysis reveals a bias in data (records and species) toward lineages associated with vertebrate hosts as well as free-living lineages. For (A) and (B) upper panels reflect the number of GenBank records while lower panels represent the number of species in GenBank with given attributes. A) The bulk of free-living species are sampled from marine systems. B) Maps reflect geographic distribution based on metadata in GenBank record, and reveals that the bulk of records do not contain country information (though the manuscripts likely do; see dataset S1, Supporting Information).

### Global distribution of SAR species with available GPS data on GenBank



### Best examples of biogeography of species based on GPS data in GenBank



**Figure 5.** Availability of biogeographic data within GenBank records. Upper panel summarizes all of the data for SAR records with GPS entries, which suggests the dominance of Rhizaria in marine areas and Alveolata and Stramenopila in terrestrial sites. Lower panel shows the global distribution of nine exemplar SAR species (three species for each clade, that had the highest number of manuscripts that contain GPS data) These insights reflect sample bias as the bulk of GenBank records have no GPS data (see Table S2 and dataset S1, Supporting Information), indicating the need for community effort to efficiently collect biogeographical data.

#### 4.2. Databases Contain Many SAR Parasite Data and Marine Free-Living Species

We next investigated the distribution of life history modes (host associated or free-living), ecology (freshwater, marine, terrestrial), and geography, based on metadata associated with GenBank entries (Figures 4 and 5). (We recognize that many more metadata exist within publications, but our goal is to assess the availability of data that can be accessed rapidly for the 21 040

species with molecular sequence on GenBank). SAR species with a host association (parasitic and other symbiotic association) represent 40% of the records and 15% of the species reported in GenBank (Figure 4). Members of Alveolata represent the majority of the host associated species (83% of the records and 86% of the species; Figure 4), almost 95% of the records and 80% of the species having a vertebrate host (e.g., *Plasmodium falciparum*, a malaria agent; *Cryptosporidium parvum*, a diarrheal agent in animals; *Babesia microti* responsible for babesiosis

known also as Texas cattle fever). A smaller proportion of Stramenopila have host-associated records and only few species are well-studied (e.g., *Blastocystis hominis*, a putative diarrheal agent in humans; *Phytophthora spp.*, plant parasites). Most Rhizaria species are not reported as host associated but instead as free-living organisms (Figure 4).

Free-living SAR lineages show the highest number of both GenBank records and species in marine systems (Figure 4). We found that one third of the records and more than half of the SAR species are missing ecological metadata (i.e., they lack isolation source, country, and habitat within GenBank entry and within the manuscript title, see methods for more details; Figure 4). This lack of metadata may indicate that ecology is not the main focus of the study (e.g., pathogenicity or molecular phylogenetic studies) or simply reflects incomplete or incorrect submission of data to GenBank as the ecology may well be in manuscripts. The number of SAR lineages characterized from freshwater and terrestrial sediments are considerably smaller than those in marine environments, suggesting that future efforts should focus on these environments in order to better understand the SAR lineages and to discover missing biodiversity.

#### 4.3. What Can Molecular Data Teach Us About SAR Biogeography?

We also evaluated the state of knowledge on biogeography of SAR species, again based on metadata associated with GenBank records. For the 4011 SAR species reported in GenBank, only 345 have GPS data (less than 9% of the species) representing a total of 2571 records with GPS data (many species have more than one GPS entry such as the foraminiferan *Pulleniatina obliquiloculata*, which have 242 GPS entries; dataset S1, Supporting Information). Fewer than 300 free-living SAR species have GPS data (the remaining 45 species are mostly parasites and the geographical information is for the non-SAR host) and less than half have a country associated in GenBank (dataset S1, Supporting Information). The scarcity of geographical data in GenBank may be related to the interest of the authors and/or to the perceived inconvenience of submitting geographical data to GenBank as they are often present in the published literature.

Despite the lack of data, we can still detect trends from the available geographical data. Records and species with geographical metadata are largely from the Northern hemisphere, particularly for Alveolata and Stramenopila, while Rhizaria appear dominant in oceans and seem to trace the course of research cruises (Figures 4 and 5). In fact, authors reporting Rhizaria and in particular, Foraminifera, have done an admirable job submitting GPS as metadata associated with sequence deposition (Figure 5). For example, the foraminiferan *Globigerinoides ruber* is present along the Morocco coast, *Globigerinella siphonifera* appears more abundant offshore, and *Globigerinoides sacculifer* lies between the two (Figure 5). Given the amount of missing data (more than half of the species did not have an entry for country (Figure 4) and less than 10% of the free-living SAR species have a GPS entry; dataset S1, Supporting Information), it is difficult to reliably make any conclusions. However, these

examples show the importance of including metadata when submitting sequences to GenBank, and suggest approaches for future data collection.

## 5. Conclusions and Outlook

Here, we show the tremendous diversity within the SAR clade and the taxa that need more work to fill gaps in knowledge for this group. Looking at the distribution of life history and ecology of the SAR species (Figure 4), we show that SAR species associated with plant hosts or more generally, non-vertebrate hosts, are understudied compared to those with a vertebrate host. The use of environmental amplicon sequencing has already increased our understanding of this dark area.<sup>[15,56,133–135]</sup> However, further work is required to provide morphological identification for environmental sequences, and to characterize lineages missed because of biases such as failure to amplify with “universal” primers (e.g., Foraminifera).<sup>[136]</sup> From a taxonomic perspective, the Rhizaria require additional study as they currently lack a clear synapomorphy and contain the bulk of undersampled families (Figures 2 and 3). Some clades require additional work because they are ecologically important groups such as the Foraminifera, which are an important link between plankton and benthos, and the Bacillariophyceae (diatoms), which includes many autotrophic species that bloom episodically. Other groups need more attention to answer fundamental questions on SAR biology (e.g., cell structures, evolution of photosynthesis) including the newly discovered taxa such as Chromerida, Bigyra, and Colponemidia. The use of an integrative taxonomic approach can provide a better understanding of species distributions including taking into account cryptic species, phenotypic plasticity, population-level variation, etc.<sup>[137–139]</sup> Recent studies have shown that the use of single cell “omics” allows acquisition of data from the dark area particularly for uncultivable species.<sup>[140,141]</sup> Hence, combining community and single cell approaches will help to increase our knowledge of this highly diverse clade.

## Supporting Information

Supporting information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors thanks members of the Katz lab, Lane lab and Delwiche lab, the two anonymous reviewers and the editors for their careful read of earlier versions of this manuscript. This work was supported by the USA National Science Foundation awards OCE-1436003 and DEB-1541511 to LAK; by the USA National Science Foundation award DEB-1541510 to CEL; by the USA National Science Foundation award DEB-1541529 and by the University of Maryland Agricultural Experiment Station to CFD.

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

biogeography ecology, eukaryotic diversity, molecular data, morphology, SAR clade (Stramenopila, Alveolata, Rhizaria)

Received: October 16, 2017

Revised: January 16, 2018

Published online: March 7, 2018

- [1] M. D. Brasier, J. Antcliffe, M. Saunders, D. Wacey, *Proc. Natl. Acad. Sci.* **2015**, *112*, 4859.
- [2] A. H. Knoll, *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, 016121.
- [3] F. Azam, T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyerreil, F. Thingstad, *Mar. Ecol. Prog. Ser.* **1983**, *10*, 257.
- [4] D. A. Caron, A. Z. Worden, P. D. Countway, E. Demir, K. B. Heidelberg, *ISME J.* **2009**, *3*, 4.
- [5] B. J. Finlay, *Environ. Microbiol.* **2002**, *296*, 1061.
- [6] S. J. Gould, *Planet of the Bacteria*. Washington Post Horizon, Washington, DC **1996**.
- [7] M. Haygood, *EcoHealth* **2010**, *7*, 400.
- [8] J. L. Ingraham, *March of the Microbes: Sighting the Unseen*. Harvard University Press, Cambridge, MA **2010**.
- [9] R. Margalef, *Our Biosphere, Vol. 10*. Ecology Institute, Oldendorfe, Luhe **1997**.
- [10] E. A. Mayer, R. Knight, S. K. Mazmanian, J. F. Cryan, K. Tillisch, *J. Neurosci.* **2014**, *34*, 15490.
- [11] A. Pineda, S.-J. Zheng, J. J. van Loon, C. M. Pieterse, M. Dicke, *Trends Plant Sci.* **2010**, *15*, 507.
- [12] J. A. Eisen, *PLoS Biol.* **2007**, *5*, 82.
- [13] W. F. Doolittle, *Sci. Am.* **2000**, *282*, 90.
- [14] L. W. Parfrey, J. Grant, Y. I. Tekle, E. Lasek-Nesselquist, H. G. Morrison, M. L. Sogin, D. J. Patterson, L. A. Katz, *Syst. Biol.* **2010**, *59*, 518.
- [15] C. de Vargas, S. Audic, N. Henry, J. Decelle, F. Mahe, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, M. Carmichael, J. Poulain, S. Romac, S. Colin, J. M. Aury, L. Bittner, S. Chaffron, M. Dunthorn, S. Engelen, O. Flegontova, L. Guidi, A. Horak, O. Jaillon, G. Lima-Mendez, J. Lukes, S. Malviya, R. Morard, M. Mulot, E. Scalco, R. Siano, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Tara Ocean, S. G. Acinas, P. Bork, C. Bowler, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, J. Raes, M. E. Sieracki, S. Speich, L. Stemann, S. Sunagawa, J. Weissenbach, P. Wincker, E. Karsenti, *Science* **2015**, *348*, 1261605.
- [16] F. Mahé, C. de Vargas, D. Bass, L. Czech, A. Stamatakis, E. Lara, D. Singer, J. Mayor, J. Bunge, S. Sernaker, *Nat. Ecol. Evol.* **2017**, *1*, 0091.
- [17] R. Massana, A. Gobet, S. Audic, D. Bass, L. Bittner, C. Boutte, A. Chambouvet, R. Christen, J. M. Claverie, J. Decelle, J. R. Dolan, M. Dunthorn, B. Edvardsen, I. Forn, D. Forster, L. Guillou, O. Jaillon, W. H. Kooistra, R. Logares, F. Mahe, F. Not, H. Ogata, J. Pawlowski, M. C. Pernice, I. Probert, S. Romac, T. Richards, S. Santini, K. Shalchian-Tabrizi, R. Siano, N. Simon, T. Stoeck, D. Vault, A. Zingone, C. de Vargas, *Environ. Microbiol.* **2015**, *17*, 4035.
- [18] M. C. Pernice, C. R. Giner, R. Logares, J. Perera-Bel, S. G. Acinas, C. M. Duarte, J. M. Gasol, R. Massana, *ISME J.* **2016**, *10*, 945.
- [19] D. Debros, I. Domaizon, J.-F. Humbert, L. Jardillier, C. Lepère, A. Oudart, N. Taib, *FEMS Microbiol. Ecol.* **2017**, *93*, fix023.
- [20] M. Simon, L. Jardillier, P. Deschamps, D. Moreira, G. Restoux, P. Bertolino, P. López-García, *Environ. Microbiol.* **2015**, *17*, 3610.
- [21] K. Hausmann, N. Hülsmann, R. Radek, *Protistology*, E. Schweizerbart'sche Verlagsbuchhandlung, **2003**.
- [22] M. A. Sleigh, *Protozoa and Other Protists*. CUP Archive, Cambridge, UK **1991**.
- [23] K. G. Grell, *Protozoology*. Springer Science & Business Media, Berlin, Germany **2013**.
- [24] O. R. Anderson, *Comparative Protozoology: Ecology, Physiology, Life History*. Springer Science & Business Media, Berlin, Germany **2013**.
- [25] J. del Campo, M. E. Sieracki, R. Molestina, P. Keeling, R. Massana, I. Ruiz-Trillo, *Trends Ecol. Evol.* **2014**, *29*, 252.
- [26] F. Burki, K. Shalchian-Tabrizi, M. Minge, A. Skjæveland, S. I. Nikolaev, K. S. Jakobsen, J. Pawlowski, *PLoS ONE* **2007**, *2*, 790.
- [27] J. Pawlowski, *Protistology* **2008**, *5*, 281.
- [28] L. A. Katz, J. R. Grant, *Syst. Biol.* **2015**, *64*, 406.
- [29] J. T. Harper, E. Waanders, P. J. Keeling, *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 487.
- [30] A. J. Gooday, *Palaos* **1994**, *9*, 14.
- [31] S. M. Adl, A. G. B. Simpson, C. E. Lane, J. Lukes, D. Bass, S. S. Bowser, M. W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. le Gall, D. H. Lynn, H. McManus, E. A. D. Mitchell, S. E. Mozley-Stanridge, L. W. Parfrey, J. Pawlowski, S. Rueckert, L. Shadwick, C. L. Schoch, A. Smirnov, F. W. Spiegel, *J. Eukaryot. Microbiol.* **2012**, *59*, 429.
- [32] T. Cavalier-Smith, *Microbiol. Rev.* **1993**, *57*, 953.
- [33] S. I. Nikolaev, C. Berney, J. F. Fahrni, I. Bolivar, S. Polet, A. P. Mylnikov, V. V. Aleshin, N. B. Petrov, J. Pawlowski, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8066.
- [34] D. J. Patterson, *Am. Nat.* **1999**, *154*, 96.
- [35] J. O. Corliss, *Prog. Protistology* **1986**, *1*, 1.
- [36] W. Foissner, D. Hawksworth, *Protist Diversity and Geographical Distribution, Vol. 8*. Springer Science & Business Media, Berlin, Germany **2009**.
- [37] M. Schaechter, *Eukaryotic Microbes*. Academic Press, San Diego, CA **2011**.
- [38] A. G. Simpson, A. J. Roger, *Curr. Biol.* **2004**, *14*, R693.
- [39] P. J. Keeling, G. Burger, D. G. Durnford, B. F. Lang, R. W. Lee, R. E. Pearlman, A. J. Roger, M. W. Gray, *Trends Ecol. Evol.* **2005**, *20*, 670.
- [40] M. Elias, J. M. Archibald, *Bioessays* **2009**, *31*, 1273.
- [41] D. Patterson, Stramenopiles: Chromophytes from a Protistan Perspective. In: *Chromophyte Algae: Problems and Perspectives*, Clarendon Press, Oxford, UK **1989**, 357.
- [42] T. Cavalier-Smith, *Progr. Phycol. Res.* **1986**, *4*, 309.
- [43] T. Cavalier-Smith, *Archiv für Protistenkunde* **1997**, *148*, 253.
- [44] R. A. Andersen, Y. Van de Peer, D. Potter, J. P. Sexton, M. Kawachi, T. Lajeunesse, *Protist* **1999**, *150*, 71.
- [45] D. Potter, G. Saunders, R. Andersen, *Am. J. Bot.* **1997**, *84*, 966.
- [46] A. Geller, D. G. Müller, *J. Exp. Biol.* **1981**, *92*, 53.
- [47] S. L. Baldauf, *J. Syst. Evol.* **2008**, *46*, 263.
- [48] S. Matsunaga, H. Uchida, M. Iseki, M. Watanabe, A. Murakami, *Photochem. Photobiol.* **2010**, *86*, 374.
- [49] G. Fu, C. Nagasato, S. Oka, J. M. Cock, T. Motomura, *Protist* **2014**, *165*, 662.
- [50] R. A. Andersen, *Am. J. Bot.* **2004**, *91*, 1508.
- [51] C. E. Hinchliff, S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazis, K. Gude, D. S. Hibbett, L. A. Katz, H. D. t. Laughinghouse, E. J. McTavish, P. E. Midford, C. L. Owen, R. H. Ree, J. A. Rees, D. E. Soltis, T. Williams, K. A. Cranston, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12764.
- [52] F. E. Round, R. M. Crawford, D. G. Mann, *Diatoms: Biology and Morphology of the Genera*. Cambridge University Press, Cambridge, UK **1990**.
- [53] T. Cavalier-Smith, J. M. Scoble, *Eur. J. Protistol.* **2013**, *49*, 328.
- [54] V. V. Aleoshin, A. P. Mylnikov, G. S. Mirzaeva, K. V. Mikhailov, S. A. Karpov, *Frontiers in Microbiol.* **2016**, *7*, 1194.

- [55] N. Yubuki, T. Pánek, A. Yabuki, I. Čepička, K. Takishita, Y. Inagaki, B. S. Leander, *J. Eukaryot. Microbiol.* **2015**, *62*, 532.
- [56] R. Massana, J. Del Campo, M. E. Sieracki, S. Audic, R. Logares, *ISME J.* **2014**, *8*, 854.
- [57] R. S. Roy, D. C. Price, A. Schliep, G. Cai, A. Korobeynikov, H. S. Yoon, E. C. Yang, D. Bhattacharya, *Sci. Rep.* **2014**, *4*, 4780.
- [58] T. Shiratori, T. Nakayama, K.-I. Ishida, *Protist* **2015**, *166*, 337.
- [59] R. FioRito, C. Leander, B. Leander, *Marine Biol.* **2016**, *163*, 1.
- [60] E. Kim, B. Sprung, S. Duhamel, C. Filardi, M. Kyoong Shin, *Environ. Microbiol.* **2016**, *18*, 4549.
- [61] J. W. Brown, U. Sorhannus, *PLoS ONE* **2010**, *5*, 12759.
- [62] R. Derelle, P. López-García, H. Timpano, D. Moreira, *Mol. Biol. Evol.* **2016**, *33*, 2890.
- [63] N. Yubuki, B. S. Leander, J. D. Silberman, *Protist* **2010**, *161*, 264.
- [64] R. Massana, J. Castresana, V. Balague, L. Guillou, K. Romari, A. Groisillier, K. Valentin, C. Pedros-Alio, *Appl. Environ. Microb.* **2004**, *70*, 3528.
- [65] R. Massana, F. Unrein, R. Rodríguez-Martínez, I. Forn, T. Lefort, J. Pinhassi, F. Not, *ISME J.* **2009**, *3*, 588.
- [66] T. A. Richards, D. Bass, *Curr. Opin. Microbiol.* **2005**, *8*, 240.
- [67] T. A. Richards, A. A. Veprikitskiy, D. E. Gouliamova, S. A. Nierzwicki-Bauer, *Environ. Microbiol.* **2005**, *7*, 1413.
- [68] H. Yoon, R. Andersen, S. Boo, D. Bhattacharya, *Stramenopiles. In: Eukaryotic Microbes*, Academic Press, San Diego, CA **2011**.
- [69] K. Weckström, J. Weckström, K. Huber, C. Kamenik, R. Schmidt, W. Salvenmoser, M. Rieradevall, T. Weisse, R. Psenner, R. Kurmayer, *Arctic, Antarctic, and Alpine Res.* **2016**, *48*, 361.
- [70] J. Baier, J. F. Negendank, B. Zolitschka, *The Climate in Historical Times*. Springer, Berlin Heidelberg, Germany **2004**, p. 195.
- [71] T. Cavalier-Smith, *The Biology of Free-Living Heterotrophic Flagellates* **1991**, *113*, 131.
- [72] A. A. Gajadhar, W. C. Marquardt, R. Hall, J. Gunderson, E. V. Ariztia-Carmona, M. L. Sogin, *Mol. Biochem. Parasitol.* **1991**, *45*, 147.
- [73] M. Schlegel, *Eur. J. Protistol.* **1991**, *27*, 207.
- [74] M. L. Sogin, *Am. Zool.* **1989**, *29*, 487.
- [75] E. Haeckel, *Jenaische Zeitschrift für Medizin und Naturwissenschaft* **1873**, *7*, 561.
- [76] F. Stein, *Der Organismus der Infusionsthiere Nach Eigenen Forschungen in Systematischer Reihenfolge bearbeitet*, Vol. 2., Engelmann, **1867**.
- [77] D. M. Prescott, *Microbiol. Rev.* **1994**, *58*, 233.
- [78] G. Fang, T. R. Cech, *J. Cell Biol.* **1995**, *130*, 243.
- [79] K. Kruger, P. J. Grabowski, A. J. Zaugg, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147.
- [80] E. Meyer, D. L. Chalker, *Epigenetics of Ciliates. In: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY* **2007**, 127.
- [81] L. W. Parfrey, D. J. Lahr, L. A. Katz, *Mol. Biol. Evol.* **2008**, *25*, 787.
- [82] M. Nowacki, L. F. Landweber, *Curr. Opin. Microbiol.* **2009**, *12*, 638.
- [83] O. A. Pilling, A. J. Rogers, B. Gulla-Devaney, L. A. Katz, *Eur. J. Protistol.* **2017**, *61*, 366.
- [84] M. A. Sleight, *eLS* **2006**.
- [85] F. J. R. Taylor, *The Biology of Dinoflagellates*, Vol. 21. Blackwell Scientific Publications, Oxford UK **1987**.
- [86] D. M. Anderson, P. Hoagland, Y. Kaoru, A. W. White, *Estimated Annual Economic Impacts From Harmful Algal Blooms (HABs) in the United States*. Woods Hole Oceanographic Institution, Woods Hole, MA **2000**.
- [87] M. A. Coffroth, C. F. Lewis, S. R. Santos, J. L. Weaver, *Curr. Biol.* **2006**, *16*, R985.
- [88] T. C. Lajeunesse, *Marine Biol.* **2002**, *141*, 387.
- [89] F. Livolant, Y. Bouligand, *Chromosoma* **1978**, *68*, 21.
- [90] P. J. Rizzo, L. D. Noodén, *Science* **1972**, *176*, 796.
- [91] G. I. McFadden, M. E. Reith, J. Munholland, N. Lang-Unnasch, *Nature* **1996**, *381*, 482.
- [92] S. Köhler, C. F. Delwiche, P. W. Denny, L. G. Tilney, P. Webster, R. J. M. Wilson, J. D. Palmer, D. S. Roos, *Science* **1997**, *275*, 1485.
- [93] R. B. Moore, M. Oborník, J. Janouškovec, T. Chrudimský, M. Vancová, D. H. Green, S. W. Wright, N. W. Davies, C. J. S. Bolch, K. Heimann, J. Šlapeta, O. Hoegh-Guldberg, J. M. Logsdon, D. A. Carter, *Nature* **2008**, *451*, 959.
- [94] M. Oborník, J. Lukeš, *Annu. Rev. Microbiol.* **2015**, *69*, 129.
- [95] T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 297.
- [96] F. Burki, P. J. Keeling, *Curr. Biol.* **2014**, *24*, R103.
- [97] P. Gilson, G. I. McFadden, *Chromosoma* **1995**, *103*, 635.
- [98] P. R. Gilson, G. I. McFadden, *Bioessays* **1997**, *19*, 167.
- [99] B. Marin, E. C. Nowack, M. Melkonian, *Protist* **2005**, *156*, 425.
- [100] F. Burki, J. Pawlowski, *Mol. Biol. Evol.* **2006**, *23*, 1922.
- [101] F. Burki, K. Shalchian-Tabrizi, J. Pawlowski, *Biol. Lett.* **2008**, *4*, 366.
- [102] J. D. Hackett, H. S. Yoon, S. Li, A. Reyes-Prieto, S. E. Rummele, D. Bhattacharya, *Mol. Biol. Evol.* **2007**, *24*, 1702.
- [103] D. Bhattacharya, T. Helmchen, M. Melkonian, *J. Eukaryot. Microbiol.* **1995**, *42*, 65.
- [104] E. Haeckel, Report of the Voyage of HMS Challenger, *Zoology* **1887**, *18*, 1.
- [105] T. Biard, L. Stemann, M. Picheral, N. Mayot, P. Vandromme, H. Hauss, G. Gorsky, L. Guidi, R. Kiko, F. Not, *Nature* **2016**, *532*, 504.
- [106] L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, L. Stemann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. de Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, C. Tara Oceans Consortium *Nature* **2016**, *532*, 465.
- [107] A. T. Howe, D. Bass, K. Vickerman, E. E. Chao, T. Cavalier-Smith, *Protist* **2009**, *160*, 159.
- [108] C. B. Harder, R. Ronn, A. Brejnrod, D. Bass, W. A. Al-Soud, F. Ekelund, *ISME J.* **2016**, *10*, 2488.
- [109] S. T. Bates, J. C. Clemente, G. E. Flores, W. A. Walters, L. W. Parfrey, R. Knight, N. Fierer, *ISME J.* **2013**, *7*, 652.
- [110] S. Geisen, A. T. Tveit, I. M. Clark, A. Richter, M. M. Svenning, M. Bonkowski, T. Urich, *ISME J.* **2015**, *9*, 2178.
- [111] V. E. Jassey, C. Signarbieux, S. Hattenschwiler, L. Bragazza, A. Buttler, F. Delarue, B. Fournier, D. Gilbert, F. Laggoun-Defarge, E. Lara, R. T. Miles, E. A. Mitchell, R. J. Payne, B. J. Robroek, *Sci. Rep.* **2015**, *5*, 16931.
- [112] J. Pawlowski, *Micropaleontology* **2000**, *46*, 1.
- [113] J. M. Archibald, *IUBMB Life* **2005**, *57*, 539.
- [114] A. Bodył, *J. Phycol.* **2005**, *41*, 712.
- [115] N. J. Patron, Y. Inagaki, P. J. Keeling, *Curr. Biol.* **2007**, *17*, 887.
- [116] M. Sanchez-Puerta, C. F. Delwiche, *J. Phycol.* **2008**, *44*, 1097.
- [117] J. M. Archibald, *Curr. Biol.* **2009**, *19*, R81.
- [118] P. J. Keeling, *J. Eukaryot. Microbiol.* **2009**, *56*, 1.
- [119] P. J. Keeling, *Phil. Trans. R. Soc. B: Biol. Sci.* **2010**, *365*, 729.
- [120] A. Bodył, J. W. Stiller, P. Mackiewicz, *Trends Ecol. Evol.* **2009**, *24*, 119.
- [121] A. Bodył, *Biol. Rev.* **2018**, *93*, 201.
- [122] J. Janouškovec, A. Horák, M. Oborník, J. Lukeš, P. J. Keeling, *Proc. Natl. Acad. Sci.* **2010**, *107*, 10949.
- [123] W.-T. Liu, T. L. Marsh, H. Cheng, L. J. Forney, *Appl. Environ. Microb.* **1997**, *63*, 4516.
- [124] F. V. Wintzingerode, U. B. Göbel, E. Stackebrandt, *FEMS Microbiol. Rev.* **1997**, *21*, 213.
- [125] C. A. Lupone, R. Knight, *Proc. Natl. Acad. Sci.* **2007**, *104*, 11436.
- [126] J.-J. Godon, E. Zumstein, P. Dabert, F. Habouzit, R. Moletta, *Appl. Environ. Microb.* **1997**, *63*, 2802.
- [127] L. Guillou, W. Eikrem, M.-J. Chrétiennot-Dinet, F. Le Gall, R. Massana, K. Romari, C. Pedros-Alió, D. Vaultot, *Protist* **2004**, *155*, 193.
- [128] S. M. Huse, L. Dethlefsen, J. A. Huber, D. M. Welch, D. A. Relman, M. L. Sogin, *PLoS Genet.* **2008**, *4*, 1000255.

- [129] N. R. Pace, *Science* **1997**, 276, 734.
- [130] T. Cavalier-Smith, *Biol. Rev. Cambridge Philosophic. Soc.* **1998**, 73, 203.
- [131] D. Bass, A. T. Howe, A. P. Mylnikov, K. Vickerman, E. E. Chao, J. Edwards Smallbone, J. Snell, C. Cabral, Jr., T. Cavalier-Smith, *Protist* **2009**, 160, 483.
- [132] D. J. G. Lahr, H. D. I. Laughinghouse, F. Gao, A. Oliverio, L. A. Katz, *Bioessays* **2014**, 36, 950.
- [133] J. D. Grattepanche, L. F. Santoferrara, G. B. McManus, L. A. Katz, *Mol. Ecol.* **2016**, 25, 3987.
- [134] S. Monchy, J. D. Grattepanche, E. Breton, D. Meloni, G. Sanciu, M. Chabe, L. Delhaes, E. Viscogliosi, T. Sime-Ngando, U. Christaki, *PLoS One* **2012**, 7, e39924.
- [135] L. F. Santoferrara, J. D. Grattepanche, L. A. Katz, G. B. McManus, *ISME J.* **2016**, 10, 1779.
- [136] J. D. Grattepanche, L. F. Santoferrara, G. B. McManus, L. A. Katz, *Trends Microbiol.* **2014**, 22, 432.
- [137] P. Z. Goldstein, R. DeSalle, *Bioessays* **2011**, 33, 135.
- [138] J. M. Padiá, A. Miralles, I. De la Riva, M. Vences, *Front. Zool.* **2010**, 7, 16.
- [139] K. W. Will, B. D. Mishler, Q. D. Wheeler, *Syst. Biol.* **2005**, 54, 844.
- [140] Y. I. Tekle, F. C. Wood, L. A. Katz, M. A. Ceron-Romero, L. A. Gorfú, *Genome. Biol. Evol.* **2017**, 9, 375.
- [141] X. Maurer-Alcalá, R. Knight, L. A. Katz, *mBio*, **2018**, 9, e01836.