MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 4328–4338 doi: 10.1111/j.1365-294X.2010.04813.x

Complex phylogeographic patterns in the freshwater alga *Synura* provide new insights into ubiquity vs. endemism in microbial eukaryotes

SUNG MIN BOO,* HAN SOON KIM,† WOONGGHI SHIN,* GA HUN BOO,* SUNG MI CHO,* BOK YEON JO,* JEE-HWAN KIM,* JIN HEE KIM,† EUN CHAN YANG,‡ PETER A. SIVER,§ ALEXANDER P. WOLFE,¶ DEBASHISH BHATTACHARYA,** ROBERT A. ANDERSEN‡ and HWAN SU YOON‡

**Department of Biology, Chungnam National University, Daejeon 305-764, Korea, †Department of Biology, Kyungbook National University, Daegu 702-701, Korea, ‡Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA, §Department of Botany, Connecticut College, New London, CT 06320, USA, ¶Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB T6G 2E3, Canada, **Department of Ecology, Evolution and Natural Resources and Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA*

Abstract

The global distribution, abundance, and diversity of microscopic freshwater algae demonstrate an ability to overcome significant barriers such as dry land and oceans by exploiting a range of biotic and abiotic colonization vectors. If these vectors are considered unlimited and colonization occurs in proportion to population size, then globally ubiquitous distributions are predicted to arise. This model contrasts with observations that many freshwater microalgal taxa possess true biogeographies. Here, using a concatenated multigene data set, we study the phylogeography of the freshwater heterokont alga *Synura petersenii sensu lato*. Our results suggest that this *Synura* morphotaxon contains both cosmopolitan and regionally endemic cryptic species, co-occurring in some cases, and masked by a common ultrastructural morphology. Phylogenies based on both proteins (seven protein-coding plastid and mitochondrial genes) and DNA (nine genes including ITS and 18S rDNA) reveal pronounced biogeographic delineations within phylotypes of this cryptic species complex while retaining one clade that is globally distributed. Relaxed molecular clock calculations, constrained by fossil records, suggest that the genus *Synura* is considerably older than currently proposed. The availability of tectonically relevant geological time (10^7–10^8 years) has enabled the development of the observed, complex biogeographic patterns. Our comprehensive analysis of freshwater algal biogeography suggests that neither ubiquity nor endemism wholly explains global patterns of microbial eukaryote distribution and that processes of dispersal remain poorly understood.

Keywords: biogeography, endemism, *Synura*, Synurophyceae, ubiquity hypothesis

Received 3 May 2010; revised received 26 July 2010, accepted 30 July 2010

Introduction

Freshwater microscopic organisms must overcome significant barriers to become distributed over vast geographic expanses. Birds, insects, and wind have been suggested as potential vectors to transport viable algal cells, enabling dispersal across ocean basins and continental land masses (Messikommer 1943; Schlichting 1961). These colonization vectors have been considered unlimited, leading to the hypothesis that free-living microbial eukaryotes <1 mm in size are distributed globally in proportion to the size of their populations and therefore lack biogeographies (Finlay 2002). This

Correspondence: Hwan Su Yoon, Fax: 207 633 9641;
E-mail: hsyoon@bigelow.org

model of microbial distribution is known as the ubiquity hypothesis, for which the metaphor 'everything is everywhere, the environment selects' has arisen (Beijerinck 1913; Baas Becking 1934; Finlay & Clarke 1999; Finlay 2002; Finlay *et al.* 2002). Among freshwater protists, the observed distribution of some taxa, particularly morphospecies of ciliates, appears to lend support to the ubiquity hypothesis (Finlay & Clarke 1999). However, evidence of dispersal limitation among other taxonomic groups suggests that this model may not be universally applicable to all microorganisms (LaChance 2004). For example, phylospecies of microscopic fungi exhibit pronounced biogeographic delineations, which are not always morphologically apparent (Taylor *et al.* 2006). The widespread presence of such cryptic species complexes, as revealed by phylogeography, may simply reflect the fact that single-celled microorganisms often have insufficient diagnostic morphological characters and developmental stages relative to multicellular plants and animals (Foissner 2008).

With respect to freshwater microalgae, diatom metacommunities reveal pronounced interhemispheric diversity gradients, suggesting poor historical dispersal to the high southern latitudes (Vyverman *et al.* 2007). In Europe and North America, freshwater diatom distribution is highly provincial rather than ubiquitous, because richness optima are associated with environmental optima that are geographically heterogeneous (Telford *et al.* 2006). Regardless of emerging evidence suggesting that many microbial eukaryotes possess true biogeographies, the ubiquity hypothesis applies in certain cases, for example the open ocean where dispersal may indeed be unlimited (Cermeño & Falkowski 2009). Clearly, biogeographic studies of microbial eukaryotes are hampered when species concepts are based solely on morphology, leading to potential incongruence between taxonomy and phylogeny. Furthermore, the insufficience of isolates from widespread field collections limits the capacity to test rigorously hypotheses concerning microbial distribution.

Synurophyceae (Phylum Heterokontophyta) are common freshwater flagellates known from all continents except Antarctica (Kristiansen 2008). They occur principally in dilute and circumneutral to acidic lakes and ponds. Members of the Synurophyceae produce ornamented siliceous components, including bristles, cysts, and scales that are taxonomically diagnostic and have considerable preservation potential in the fossil record. The siliceous remains of the Synurophyceae enable the concept of morphospecies, defined here as groups of individuals or populations with the same or similar morphological characters (Evans *et al.* 2007), to be applied reproducibly. However, it must also be assumed that ultrastructure alone may betray genetic differences within morphospecies, as witnessed in other groups for which molecular data have confirmed the presence of cryptic species complexes (Blackburn & Tyler 1987; Montresor *et al.* 2003; Darling *et al.* 2004; Sarno *et al.* 2005).

Current synurophycean taxonomic schemes are based primarily on morphological examinations using scanning electron microscopy (SEM), which reveal morphospecies that are both cosmopolitan as well as regionally endemic (Kristiansen 1996, 2008; Vigna & Siver 2003; Kristiansen & Lind 2004). However, the apparent cosmopolitan nature of some taxa has yet to be conclusively demonstrated by the integration of genetic, morphological, and palaeontological approaches. In this study, we investigate the extent of cryptic diversity within the cosmopolitan morphospecies *Synura petersenii sensu lato*. This work was stimulated by the initial discovery of unsuspected differentiation within ITS sequences from natural populations from South Korea. We subsequently surveyed all known culture isolates of this morphospecies to assess global phylogeographic patterns using conserved protein markers. We then conducted molecular clock analyses, constrained by various dates from the fossil record, and compared these results *a posteriori* to observations of fossil scaled synurophytes from Middle Eocene lake sediments. The existence of genetically distinct phylospecies within the *Synura petersenii* morphospecies suggests that cryptic species may have both global and provincial ranges of distribution, challenging the polarization of current hypotheses concerning ubiquity and endemism.

Materials and methods

Samples and cell culture

The 63 strains of *Synura petersenii* were isolated from 63 different reservoirs in South Korea. Details of collection sites are provided in Table S1. All the isolates were grown in DY III medium (Lehman 1976) buffered to pH 7 at the culture collections at Chungnam National University (CNU) and Kyungbuk National University. The unialgal cultures were maintained at 15 ± 1 °C with an illumination of 100 μmol/m^2 per s with cool white fluorescent light at a 14:10 h light–dark cycle. In addition, 33 strains outside South Korea were obtained from the following algal culture collections (Table S1): Coimbra Collection of Algae (ACOI), Portugal (http://www.acoi.ci.uc.pt); Culture Collection of Algae at the University of Cologne (CCAC), Germany (http://www.ccac.uni-koeln.de/); Culture Collection of Algae and Protozoa, (CCAP) England (http://www.ccap.ac.uk/); The Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP), United States (http://ccmp.

bigelow.org/); Microbial Culture Collection at National Institute for Environmental Studies (MCC-NIES), Japan (http://www.mcc.nies.go.jp/); Culture Collection of Algae at the University of Göttingen (SAG), Germany (http://www.epsag.uni-goettingen.de/); The Culture Collection of Algae at The University of Texas at Austin (UTEX), United States (http://www.utex.org/). The outgroup taxa were *S. uvella*, *Synura* sp., and *Mallomonas caudata* (collected in South Korea), *S. curtispina* and *S. echinulata* (from SAG), and *S. sphagnicola*, *Hibberdia magna*, *Ochromonas tuberculata*, and *Chromulina nebulosa* (from CCMP).

DNA purification, PCR amplification, and sequencing

Live cells in a culture tube of each strain were used for DNA extraction. Extraction buffer was added directly into tubes of the concentrated, harvested cells and DNA extracted as described in the Invisorb Spin Plant Mini Kit (Invitek, Berlin-Buch, Germany). Detailed primer information is listed in Table S2. Polymerase chain reactions (PCRs) were carried out in 25 μL of reaction cocktail containing DNA template (1–10 ng), 10X *Ex Taq*™ buffer (Mg^{2+} free), 25 mM $MgCl_2$, dNTP (2.5 mM each), and 10 pmol of each primer. Amplification was performed using proofreading *TaKaRa Ex Taq*™ DNA polymerase (TaKaRa Shuzo Co., Tokyo, Japan) for an initial denaturation step at 95 °C for 4 min, followed by 30 cycles (35 cycles for *rbc*L and *cox*1 region) with a denaturing step at 95 °C for 30 s (1 min for *rbc*L and *cox*1), a primer annealing step at 50–55 °C for 30 s (1 min for *rbc*L and *cox*1), an extension step at 72 °C for 1 min (2 min for *rbc*L and *cox*1). The amplification was terminated with a final extension at 72 °C for 6 min. PCR products were purified using the High Pure™ PCR Product Purification Kit (Roche, Indianapolis, IN, USA) following the manufacturer's instructions. Purified PCR products were sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA). All sequences of the forward and reverse strands were determined for all taxa using from the ABI Prism™ 377 DNA Sequenator (Applied Biosystems, Foster City, CA, USA) at Research Center, Chungnam National University, Daejeon, Korea, and ABI 3730xl DNA Analyzer (Applied Biosystems). The electropherogram output for each specimen was edited using the program SEQUENCE NAVIGATOR v. 1.0.1 (Applied Biosystems) and rechecked manually for consistency.

Phylogenetic reconstruction

Nucleotide sequences were aligned using SE-AL v2.0a11 (Sequence Alignment Editor Version 2.0 alpha 11; http://tree.bio.ed.ac.uk/software/seal/). Hypervariable regions in rDNA were realigned with relaxed gap opening and extensions and manually verified. Phylogenies using individual genes and 9-gene concatenated DNA alignment were reconstructed by maximum-likelihood (ML) method with the GTR + Γ nucleotide model using RAXML v.7.0.4 (Stamatakis 2006). The evolutionary models were chosen on the BIC criterion basis implemented in MODELGENERATOR version 0.85 (Keane *et al.* 2006). Aligned nucleotide sequences excluded ITS and 18S rDNA were translated into amino acid sequences using the MACCLADE program (Maddison & Maddison 2002). A phylogeny using the combined 7-gene amino acid sequence was inferred with the cpREV + Γ evolutionary model using RAXML. We used 200 independent tree inferences using the -# option with the default -i (automatically optimized SPR rearrangement) and -c (25 distinct rate categories) options of the program to identify the best tree. Statistical support for nodes in the trees was obtained from bootstrap replications. The ML bootstrap values were calculated with 1000 replicates using the same substitution model using RAXML. One-thousand MP bootstrap replicates were carried out using PAUP* 4.0b10 (Swofford 2002). Phylogenies using individual genes and 9-gene data set are shown in the Supporting Information (Figs S1–S10).

Molecular clock analyses

The sequence alignment used for the molecular divergence analysis comprised a concatenated protein-coding plastid *psa*A, *psb*A, *psb*C, and *rbc*L data set (36 taxa 1198 aa), which is suitable for addressing ancient splits within red algae and Stramenopiles (Yoon *et al.* 2002; Cho *et al.* 2004). However, the slow-evolving SSU rDNA and hypervariable DNA sequences of ITS, *cob*, *cox*1, and *cox*3 were excluded from this analysis. The ML tree was used to estimate the divergence dates. The hypothesis of rate constancy among taxa was tested by a comparison of likelihoods, using a likelihood ratio test (Felsenstein 1988), given the best ML tree topology with and without the constraint of a molecular clock. The Bayesian molecular dating was carried out using the MULTIDIVTIME software package (Thorne & Kishino 2002). Divergence times were estimated under supergene local clock method (Hedges & Kumar 2003).

The five following constraints were used for the divergence time estimation (see Fig. 1). The first constraint was 1174–1222 million years ago (Ma) based on the red algal fossil, *Bangiomorpha* (Butterfield 2000). The second constraint was the upper limit date of 603 Ma based on the presence of corallinalean red algae from deposits in China (Xiao *et al.* 2004). The three other calibration points were 433–490 Ma for the origin of land plants

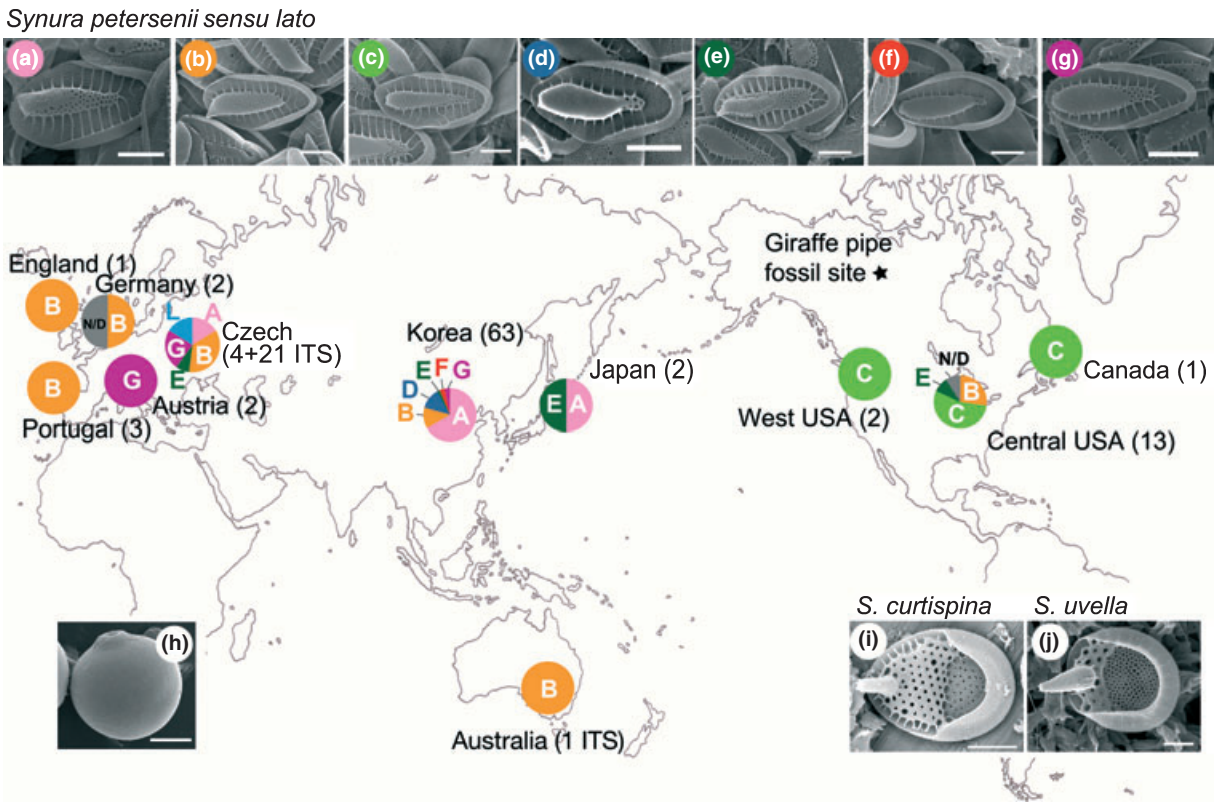


Fig. 1 Biogeographic distribution of the eight clades of *Synua petersenii* (a–g) that were identified using multigene phylogenetic analysis. Pie-charts refer to proportions of a given clade in a given geographic region. SEM micrographs of representative specimens from each clade in Fig. 2 illustrate their morphological similarity and assignment to *S. petersenii*. We used SEM to verify that all isolates conform to this morphospecies. N/D represents 'not designated' isolates. Additional micrographs include the cyst of *S. petersenii* (h) and scales of the closely related congeneric taxa, *S. curtispina* (i) and *S. uvella* (j). Scale bars are 1 μm except (h), where it is 5 μm.

(Kenrick & Crane 1997), 290–320 Ma for the angiosperm–gymnosperm split (Bowe *et al.* 2000), and 144–206 Ma for the monocot–dicot split (Sanderson & Doyle 2001). To account for the uncertainty of the molecular calibration points, we used intervals as lower and upper constraints (L and U options of constraints). The prior time from the tips to the root was set at 1500 Ma with a standard deviation of 5 Ma (i.e. rttm = 1.5, rttmsd = 0.5) to correspond with the results of Yoon *et al.* (2004). The MCMC analyses were run for one million generations with sampling each 100 generations after a burn-in point of one million generations (numsamps = 10 000, sampfreq = 100, burnin = 1 000 000). Convergence of the MCMC algorithm was assessed by running multiple analyses (more than two), each starting with a different randomly selected initial state.

Preparation of fossil material

Fossil *Synura* specimens originate from the Giraffe kimberlite fossil locality in northwestern Canada (64º44′ N, 109º45′ W). Organic mudstone samples were prepared for examination using samples archived in the laboratories of PAS and APW (Siver & Wolfe 2005a, 2009). Mudstones with *S. recurvata* and *S. curtispina* were taken from a depth of 124 m in Giraffe core 99-01, and those of taxa related to *S. petersenii* and extinct *Synura* sp. were taken from 107–109 m. Approximately 1 g of organic mudstone was oxidized with 30% H_2O_2 for 24 h, centrifuged, and rinsed several times with distilled water. Aliquots of cleaned slurry were air-dried onto heavy-duty aluminium foil, trimmed, and mounted to aluminium stubs with Apiezon® wax. The stubs were coated with an Au–Pd mixture for 2 min in a Polaron Model E sputter-coater and examined with a Leo 982 field-emission SEM.

ITS phylogeny

The initial survey of genetic diversity within *S. petersenii sensu lato* was restricted to ITS sequences from isolates originating from 63 South Korean water bodies spanning a wide range of habitats. These analyses suggested the presence of multiple cryptic (phylo)species

within the Korean *S. petersenii* morphospecies complex (Fig. S1). The relatively large number of isolates that were investigated reduces (but may not eliminate) the potential impact of sampling bias in ascertaining native genetic diversity. This stimulated the expansion of sampling to include an additional 33 cultures collected from around the world and the application of a wider range of sequences to test the initial ITS results. For this purpose, we took a conservative approach and determined the sequence of 7 highly constrained organelle-encoded markers (plastid *psa*A, *psb*A, *psb*C, *rbc*L; mitochondrial *cob, cox*1, *cox*3) from all 96 samples and generated a 7-protein alignment for phylogenetic analysis. For the ITS phylogeny, we added published sequences from isolates originating from the Czech Republic (21) and Australia (1) (see Fig. S1). Our working hypothesis was that if ITS-based phylospecies were robustly supported by the protein data, then we would have strong evidence from independent markers for relatively ancient splits and complex evolutionary trajectories. However, if the protein data provide evidence of paraphyly for the isolates, then differentiation using ITS alone may not record fixed differences or may provide spurious evolutionary signal. Plastid markers have a long history in the study of plant and algal evolution (e.g., Evans *et al*. 2007; Duangjai *et al*. 2009) and are not expected to provide specious phylogenetic signals.

Results and discussion

Phylogeography

Maximum-likelihood analysis of the concatenated protein data provides unambiguous evidence for the existence of multiple distinct phylospecies within the *S. petersenii* morphotaxon. This tree reveals seven distinct clades (A–G) that have moderate to strong bootstrap support (Figs 1 and 2). One isolate from Germany (CCAC 0052) and one from the USA (CCMP 869) were not recovered as a monophyletic lineage in the concatenated protein tree, although this group had moderate bootstrap support in the 9-gene DNA data set (68% in ML and 65% in MP) (Fig. S2). Silica scale morphology of members from each clade is very similar, although not identical, but nonetheless remains entirely consistent with designation as *S. petersenii* (Fig. 1). Clade A occurred in Korea and Japan (but see below), Clade B was found in four continents, Clade C was present in North America, Clade D and F were found in Korea, Clade E occurred in eastern Asia and North America, and Clade G was found in Korea and Europe (Fig. 2). The results are congruent with the 9-gene DNA phylogeny (Fig. S2), which includes ITS and 18S rDNA sequences, as well as the individual gene trees (Figs S3–S10). ITS sequences from the 21 Czech isolates were distributed alternatively within Clades A (Korea and Japan), B (widespread), and G (Korea and Europe), as well as a new Clade (H, 4 isolates, Fig. S1) that is sister to Clade C (North America). The ITS sequence from Australia is positioned within Clade B. Based on the multigene and ITS phylogeny, the well-supported Clades A, B, E, and G occur on two or more continents, whereas Clades D and F are restricted to Asia, Clade C to North America, and apparently Clade H to the Czech Republic.

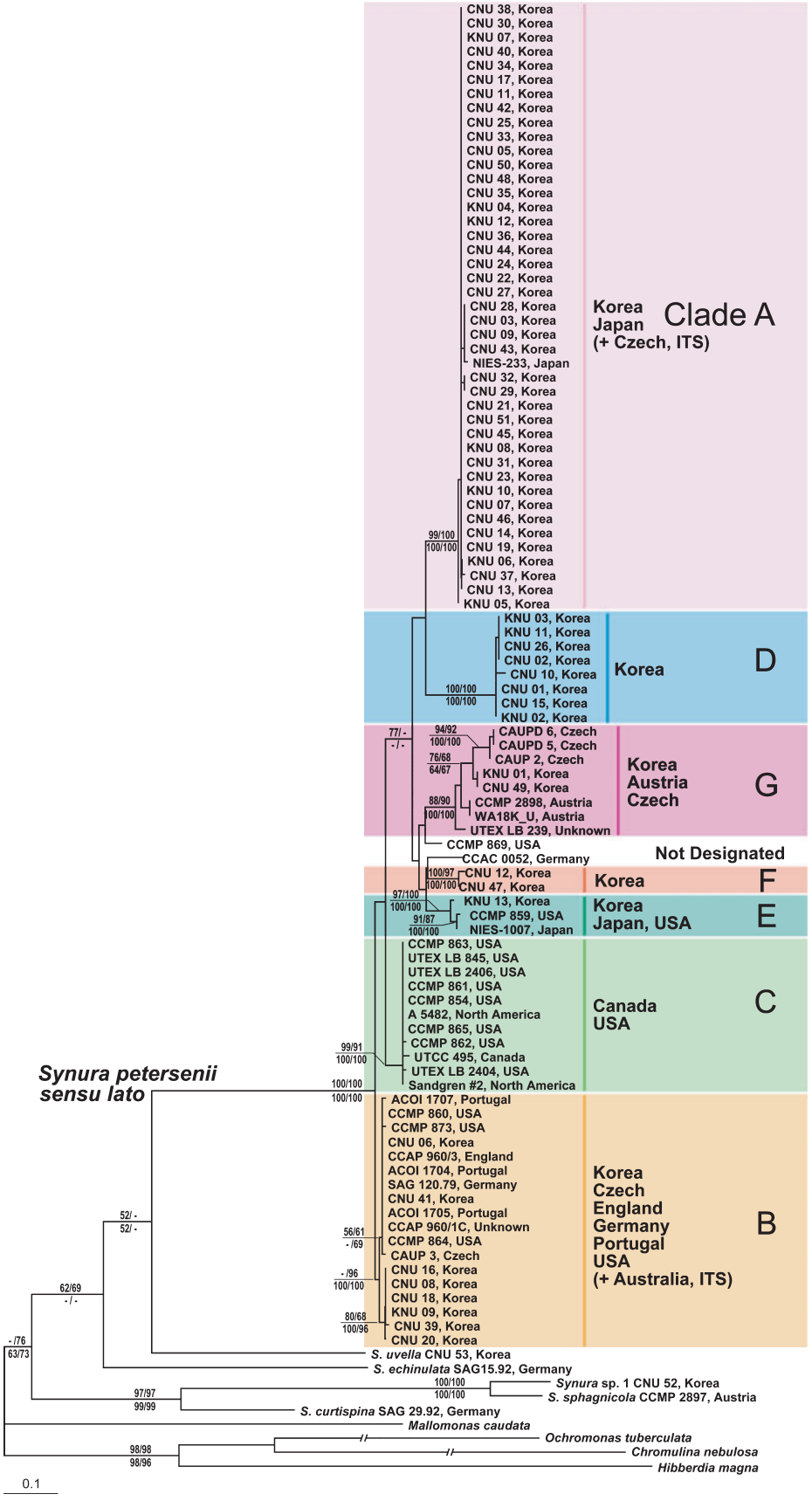
The results provide robust molecular evidence for the presence of cryptic species within the *S. petersenii* complex, of which several have restricted biogeographic distributions. Although samples from every habitat on every continent are needed to conclusively verify endemism or provincialism within any microbial phylospecies, the clear signal evident in our more limited data supports the notion that endemic species have arisen on different continents (i.e., Clades C, D, F, and H). This level of phylogenetic discrimination is unlikely to occur by chance if we are indeed sampling randomly a single gene pool comprising globally distributed cells.

The results also demonstrate, in the case of Clade B, that at least one phylospecies is globally distributed, and this group occupies a basal position on the protein ML tree (Fig. 2). Furthermore, interclade relationships suggest that the putative Korean endemics (Clades D and F) originated through genetic isolation of individuals represented by the globally distributed Clade B even though this hypothesis has only moderate bootstrap support (77%).

Limitations to dispersal

In the model of ubiquitous microbial distribution, the success of global colonization is assured by large initial populations and small cell sizes (Finlay 2002). The *S. petersenii* morphospecies clearly meets these criteria:

Fig. 2 Phylogeny of the *Synura petersenii* complex inferred from the phylogenetic analysis of a concatenated 7-protein alignment (plastid *psa*A, *psb*A, *psb*C, *rbc*L; mitochondrion *cob*, *cox*1, *cox*3), with five closely related *Synura* and other heterokonts as outgroup taxa. Colour assignments for each clade match those in Fig. 1. Values shown above and below branches are bootstrap values (1000 iterations) from the 7-gene amino acid data and the 9-gene DNA data, respectively (ML/MP). Only bootstrap values >50% are shown. Branch lengths are proportional to the number of substitutions per site (see scale bar).



it blooms in countless lakes, ponds, and streams and produces copious, viable siliceous resting cysts (Fig. 1h). Colonies have diameters of ~50 μm, which is considerably smaller than the 1- mm threshold proposed for ubiquity to arise (Finlay 2002). Despite these attributes, several clades appear to be regionally restricted, suggesting that dispersal limitation occurs for some (but perhaps not all) cryptic species within the *S. petersenii* complex.

Ecologically, there are no obvious limnological features associated with these genetically defined biogeographic patterns. For example, North American strains (Clade C) thrive in highly diverse habitats ranging from acid bogs in Newfoundland, roadside ponds in New York State, acidic ponds and streams in Michigan, circumneutral to alkaline streams and lakes in Illinois and Indiana, and circumneutral lakes on San Juan Island, Washington. In Michigan and Illinois, North American representatives of Clade B were collected within a few kilometres of sites containing Clade C organisms. Similarly, Korean strains are not separated along gradients of pH or nutrients. Together, these observations indicate that the provinciality expressed among cryptic species is unlikely associated with environmental boundaries, in contrast to the patterns apparent among diatom morphospecies (Telford *et al.* 2006).

More broadly, our results necessitate a reconsideration of the traditional vectors associated with dispersal of freshwater microbes, namely wind and animals. For example, Clade B is globally distributed having successfully populated freshwater habitats on four continents, whereas other clades apparently failed to achieve such widespread distribution. Increased sampling effort might reveal wider distributions for some species. However, given the improbability that wind or animals act differentially on some cryptic species but not others, we argue that these colonization vectors are largely ineffective with respect to *S. petersenii*, be it at the regional or continental scale. If these dispersal vectors were important, the phylogeographic (Fig. 1) and phylogenetic (Fig. 2) patterns reported here would have been discernibly overprinted.

Genetic and morphological variation

We note that Clades A, B, C, and D contain isolates with very low sequence divergences within the 9-gene DNA alignment (~8000 bp), producing *p*-distances of 0.0011, 0.0016, 0.003, and 0.0005–0.002, respectively. This likely indicates either relatively recent isolation from founder populations that have dispersed on local to continental scales, or alternatively that these clades have undergone genetic sweeps that reduce their apparent divergence. The notion of genetic sweeps is particularly supported by Clade B (Asia, Europe, North America, Australia) because the phylogeny provides no evidence for isolation-by-distance dispersal processes. Conversely, Clades C and D are highly restricted in their geographic distributions (North America and South Korea, respectively) despite forming abundant populations across each of these regions, and presumably being available for dispersal. These results underscore the complexity that is inherent to microbial biogeography when appropriate scrutiny is applied to a selected lineage (Martiny *et al.* 2006). We suggest that the patterns of sequence divergence in *S. petersenii* have been mediated by a number of processes that do not act in isolation.

Fenchel and Finlay (2004) have argued that genetic variation is not necessarily taxonomically diagnostic, because 'the likelihood of finding a new sequence whenever a representative of a nominal species is isolated is very high'. Our study challenges this assertion: in the broad array of samples and the rich sequence database we have generated, to our knowledge the largest data set for freshwater heterokonts, new culture isolates can typically be inserted into one of the well-supported clades bound by nearly identical sequence alignments. This pattern is particularly well expressed in the mitochondrial and plastid genes, as well as nuclear ITS and 18S rDNA. Owing to the conserved, universal functions of these genetic markers, the most plausible conclusion is that each clade represents a distinct cryptic species.

Although microorganisms have fewer morphological characters and lack complex developmental stages, their genetic differentiation may prove analogous to multicellular plants and animals, in that genetic changes presage reproductive isolation, which in turn leads to speciation (Taylor *et al.* 2006). In this light, organisms such as *Synura* are anticipated to undergo speciation at a genetic level long before morphological distinctions become evident (Hull & Norris 2009). Detailed analyses of scale morphology reveal that the *S. petersenii* complex contains morphological variants that can be considered taxonomically distinct (Kristiansen & Preisig 2007; Němcová *et al.* 2008). This suggests that morphospecies exist within *S. petersenii sensu lato* but are only discernable with detailed SEM observations. The sequence data reported here suggest that such subtle morphological expressions are underpinned by considerably greater genetic differentiation.

Insights from palaeobiology

Because the biogeography of many organisms is associated with continental drift (e.g., Noonan & Sites 2010), we evaluated the extent of speciation in *Synura* over

tectonic timescales. To do this, divergence times of *S. petersenii* clades from congeneric taxa and other chrysophytes were estimated using a relaxed molecular clock applied to concatenated plastid protein sequences (*psa*A, *psa*A, *psa*C, and *rbc*L), and constrained by the fossil record to produce a phylogeny of the genus (Fig. 3a).

The results suggest that the *S. petersenii* complex originated ~200 Ma (Late Triassic-Early Jurassic), immediately prior to the break-up of the supercontinent Pangea, and subsequently radiated into several phylospecies by ~50 Ma (Eocene). Although the temporal range of possible solutions is large (Fig. 3a), the fossil record from the Giraffe kimberlite locality in northern Canada demonstrates that the *S. petersenii* morphospecies, possessing the complete range of morphological characters observed in modern counterparts, had evolved by the Middle Eocene (48–40 Ma, Fig. 3b–d), in excellent correspondence with the molecular clock calculations. The spectacular array of synurophycean fossils from this deposit also includes close congeneric forms with morphological affinities to extant *S. curtispina* (Fig. 3e) and *S. uvella* (Fig. 3f) [see modern *S. curtispina* and *S. uvella* in Figs 1i, j, respectively]. The Giraffe sediments thus provide an important milepost for scaled synurophyte evolution: taxa with modern

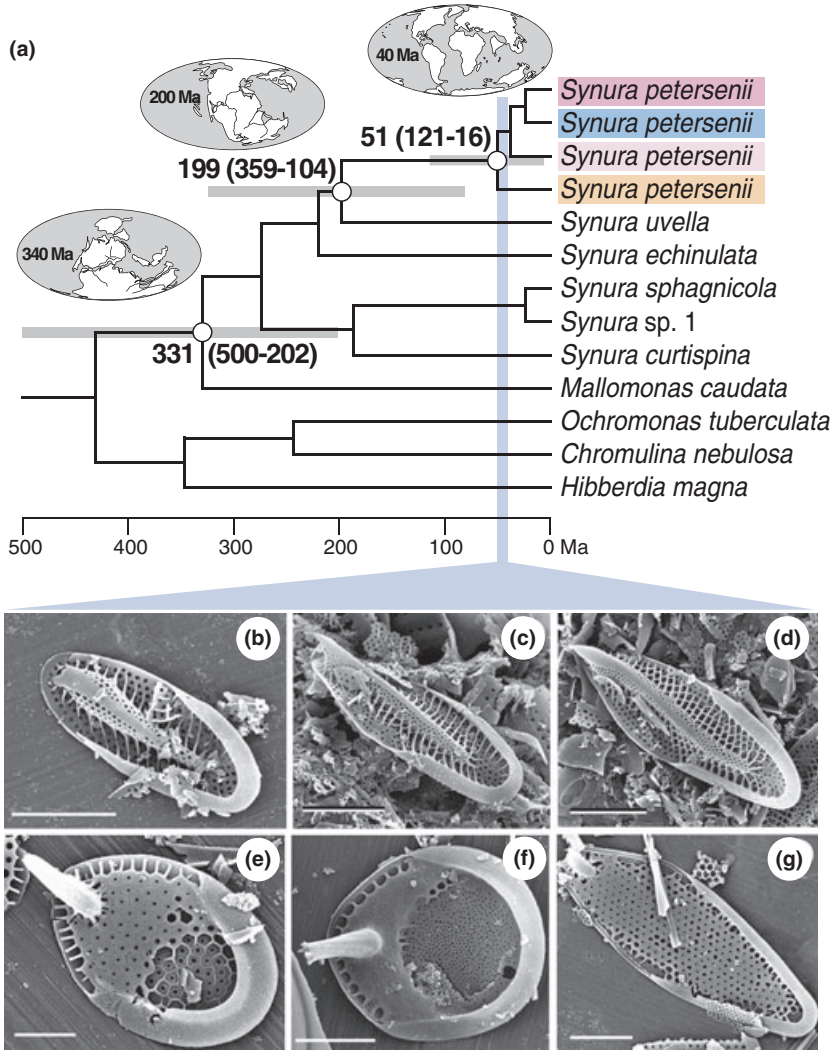


Fig. 3 Molecular divergence times of *Synura* from other chrysophytes based on *psa*A, *psb*A, *psb*C, and *rbc*L alignments and a Bayesian relaxed clock approach (a). Major evolutionary nodes (white circles) are assigned ages (in Ma) with confidence ranges (in parentheses) and corresponding continental palaeogeographic maps (as insets). The colour coding of *S. petersenii* phylospecies corresponds to that used in previous illustrations. In (b–g), SEM images of Middle Eocene fossil synurophytes illustrate taxa belonging to the *S. petersenii* morphospecies (b–d), others with affinities to modern *S. curtispina* (e) and *S. uvella* (f), and an extinct morphotaxon under study (g). Scale bars are 2 μm except (e), which is 1 μm.

affinities, including the *S. petersenii* morphospecies, occurred alongside forms that are thought to be extinct (Fig. 3g; see also Siver & Wolfe 2005b). The diversity of this Middle Eocene synurophyte community, as well as the high proportion of taxa that can be morphologically associated with modern lineages, implies protracted evolutionary stasis with regard to ultrastructure. This is once again consistent with results from the molecular clock analysis, which predict the genus *Synura* to be considerably older than currently appreciated, with an origin close to 330 Ma (Fig. 3a). The oldest unambiguous microfossils attributable to either Chrysophyceae or Synurophyceae are cysts from Lower Cretaceous marine sediments (~100 Ma) in Antarctica and Australia (Harwood & Gersonde 1990). Much older microfossils putatively interpreted as chrysophyte scales have been reported from Neoproterozoic cherts ~630 Ma (Allison & Hilgert 1986; Knoll 2003). However, these do not convincingly resemble any known chrysophyte scale, and their identity remains unclear. Irrespectively, the molecular clock calculations provide compelling evidence that geological time may be invoked to explain the complex biogeographic patterns expressed by cryptic phylospecies in the *Synura petersenii* complex. For example, ancestral forms may have been distributed by the break-up of Pangea and subsequent rifting between Europe, Greenland, and North America, yielding cosmopolitan early-divergent cryptic species. In contrast, regionally endemic phylospecies more likely originated sympatrically well after these palaeogeographic events. In the absence of effective trans-oceanic colonization vectors, these newly diverged forms were biogeographically isolated. This scenario is supported by the topology of the protein tree (Fig. 2), in which the most cosmopolitan clade is basal to those with greater provinciality.

Summary

In conclusion, our phylogeographic analysis shows that cryptic species within the *Synura petersenii sensu lato* complex can be either endemic or cosmopolitan. If this lineage is considered broadly representative of nonmarine microbial eukaryotes, then existing models of microbial distribution are not mutually exclusive of each other because phylospecies with biogeographies may arise alongside others that do not. Furthermore, the most commonly invoked vectors of microbial dispersal (wind and animals) are insufficient to explain the biogeographic patterns we have documented for the *S. petersenii* complex. Although these vectors have been deemed necessary for ubiquitous distribution to arise in freshwater algae, we also favour explanations for cosmopolitanism that reside in geological time, consistent with the results of our integrated phylogenetic and palaeobiological analysis.

Acknowledgements

This project was partially supported by grants from Korean Science and Engineering Foundation (KOSEF) to SMB (MOST; R01-2006-000-10207-0) and ECY (KRF-2008-357-C00148), from the National Science Foundation Microbial Genome Sequencing program (EF 08-27023) that was awarded to HSY, RAA, DB, from the Assembling the Tree of Life program to RAA (EF 04-31117) and to HSY and DB (DEB-0937975), and from the NSF Division of Environmental Biology grant (DEB-0716606) to PAS. APW is supported by the Natural Sciences and Engineering Research Council of Canada.

References

Allison CW, Hilgert JW (1986) Scale microfossils from the early Cambrian of northwest Canda. *Journal of Paleontology*, **60**, 973–1015.

Baas Becking LGM (1934) Geobiologie of Inleiding tot de Milieukunde. *W.P. Van Stockum & Zoon*, Den Haag, The Netherlands.

Beijerinck MW (1913) De infusies en de ontdekking der bakterien. In: *Jaarboek van de Koninklijke Akademie van Wetenschappe* (ed. Müller J). pp. 1–28, Amsterdam, The Netherlands.

Blackburn SI, Tyler PA (1987) On the nature of eclectic species – a tiered approach to genetic compatibility in the desmid *Micrasterias thomasiana*. *British Phycological Journal*, **22**, 277–298.

Bowe LM, Coat G, de Pamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 4092–4097.

Butterfield NJ (2000) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology*, **26**, 386–404.

Cermeño P, Falkowski PG (2009) Controls on diatom biogeography in the ocean. *Science*, **325**, 1539–1541.

Cho GY, Lee SH, Boo SM (2004) A new brown algal order, Ishigeales (Phaeophyceae), established on the basis of plastid protein-coding *rbc*L, *psa*A, and *psb*A region comparisons. *Journal of Phycology*, **40**, 921–936.

Darling KF, Kucera M, Pudsey CJ, Wade CM (2004) Molecular evidence links cryptic diversification in polar planktonic protists to Quaternary climate dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 7657–7662.

Duangjai S, Samuel R, Munzinger J *et al.* (2009) A multi-locus plastid phylogenetic analysis of the pantropical genus *Diospyros* (Ebenaceae), with an emphasis on the radiation and biogeographic origins of the New Caledonian endemic species. *Molecular Phylogenetics and Evolution*, **52**, 602–620.

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom ''barcode'' genes (*cox*1, *rbc*L, 18S and ITS

rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist*, **158**, 349–364.
Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics*, **22**, 521–565.
Fenchel T, Finlay BJ (2004) Here and there or everywhere? Response from Fenchel and Finlay *BioScience*, **54**, 884–885.
Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061–1063.
Finlay BJ, Clarke KJ (1999) Ubiquitous dispersal of microbial species. *Nature*, **400**, 828.
Finlay BJ, Monaghan EB, Maberly SC (2002) Hypothesis: the rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist*, **153**, 261–273.
Foissner W (2008) Protist diversity and distribution: some basic consideration. *Biodiversity and Conservation*, **17**, 235–242.
Harwood DM, Gersonde R (1990) Lower Cretaceous diatoms from ODP Leg 113 Site 693 (Wendell Sea). Part 2: Resting spores, Chrysophycean cysts, and endoskeletal dinoflagellate, and notes on the origin of diatoms. In: *Proceedings of the Ocean Drilling Program, Science Result* (eds Barker PF, Kennett JP). pp. 403–425, Ocean Drilling Program, College Station, TX.
Hedges SB, Kumar S (2003) Genomic clocks and evolutionary timescales. *Trends in Genetics*, **19**, 200–206.
Hull PM, Norris RD (2009) Evidence for abrupt speciation in a classic case of gradual evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 21224–21229.
Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology*, **6**, 29.
Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. *Nature*, **389**, 33–39.
Knoll AH (2003) Biomineralization and evolutionary history. *Reviews in Mineralogy and Geochemistry*, **54**, 329–356.
Kristiansen J (1996) *Biogeography of Freshwater Algae*, Kluwer Acad. Publishers, Dordrecht, The Netherlands.
Kristiansen J (2008) Dispersal and biogeography of silica-scaled chrysophytes. *Biodiversity and Conservation*, **17**, 419–426.
Kristiansen J, Lind JF (2004) Endemicity in silica-scaled chrysophytes. *Nova Hedwigia Beiheft*, **128**, 65–83.
Kristiansen J, Preisig HR (2007) Chrysophyte and haptophyte algae, Part 2: Synurophyceae. In: *Freshwater Flora of Central Europe* (eds Büdel B, Gärtner G, Krienitz L, Preisig HR, Schagerl M), pp. 252. Spektrum Akademischer, Verlag.
LaChance M-A (2004) Here and there or everywhere? *BioScience*, **54**, 884.
Lehman JT (1976) Ecological and nutritional studies on *Dinobryon* Ehrenb.: seasonal periodicity and the phosphate toxicity problem. *Limnology and Oceanography*, **21**, 646–658.
Maddison DR, Maddison WP (2002) *MacClade*. Sinauer, Sunderland.
Martiny JB, Bohannan BJ, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.
Messikommer E (1943) Untersuchungen über die passive Verbreitung der Algen. *Zeitschrift für Hydrologie*, **9**, 310–316.
Montresor M, Lovejoy C, Orsini L, Procaccini G, Roy S (2003) Bipolar distribution of the cyst-forming dinoflagellate *Polarella glacialis*. *Polar Biology*, **26**, 186–194.
Němcová Y, Nováková S, Řezáčová-Škaloudová M (2008) *Synura obesa* sp. nov. (Synurophyceae) and other silica-scaled chrysophytes from Abisko (Swedish Lapland). *Nova Hedwigia*, **86**, 243–254.
Noonan BP, Sites Jr JW (2010) Tracing the origins of iguanid lizards and boine snakes of the pacific. *The American Naturalist*, **175**, 61–72.
Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbc*L and 18S rDNA data. *American Journal of Botany*, **88**, 1499–1516.
Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of S. costatum-like species with the description of four new species. *Journal of Phycology*, **42**, 151–176.
Schlichting HEJ (1961) Viable species of algae and protozoa in the atmosphere. *Lloydia*, **24**, 81–88.
Siver PA, Wolfe AP (2005a) Eocene scaled chrysophytes with pronounced modern affinities. *International Journal of Plant Sciences*, **166**, 533–536.
Siver PA, Wolfe AP (2005b) Scaled chrysophytes in Middle Eocene lake sediments from Northwestern Canada, including descriptions of six new species. *Nova Hedwigia Beiheft*, **128**, 295–308.
Siver PA, Wolfe AP (2009) Tropical ochrophyte algae from the eocene of northern Canada: a biogeographic response to past global warming. *Palaios*, **24**, 192–198.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods) 4.0b8*, Sinauer, Sunderland, MA.
Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **361**, 1947–1963.
Telford RJ, Vandvik V, Birks HJ (2006) Dispersal limitations matter for microbial morphospecies. *Science*, **312**, 1015.
Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, **51**, 689–702.
Vigna MS, Siver PA (2003) Biodiversity and biogeographical implications of silica-scaled chrysophytes (Chrysophyceae and Synurophyceae) of the northeast wetlands of Argentina. *Archiv für Hydrobiologie*, **158**, 359–372.
Vyverman W, Verleyen E, Sabbe K *et al.* (2007) Historical processes constrain patterns in global diatom diversity. *Ecology*, **88**, 1924–1931.
Xiao S, Knoll AH, Yuan X, Pueschel CM (2004) Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. *American Journal of Botany*, **91**, 214–227.
Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single, ancient origin of chromist plastids. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15507–15512.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology and Evolution*, **21**, 809–818.

SMB, GHB, SMC and J-HK are studying the systematics, biogeography and evolution of algae. HSK and JHK are interested in the ecology and taxonomy of freshwater algae. WS and BYJ are working on a molecular and ultrastructural study of freshwater algae. PAS studies the taxonomy, distribution, ecology and paleobiology of algae with an emphasis on siliceous forms. APW investigates the paleobiology of global environmental change. DB works on the evolutionary and functional genomics of algae and protists. RAA studies the systematics of algae. HSY and ECY are working on the systematics and genomics of algae.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information of samples used in the present study with 896 new sequences

Fig. S1 ITS RAxML phylogenetic tree of *Synura petersenii sensu lato*, including published sequences from the Czech Republic (blue) and Australia (red).

Fig. S2 RAxML phylogenetic tree using the 9-gene DNA concatenated data set.

Fig. S3 SSU rDNA RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S4 *psa*A RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S5 *psb*A RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S6 *psb*C RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S7 *rbc*L RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S8 *cox*1 RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S9 *cox*3 RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Fig. S10 *cob* RAxML phylogenetic tree of *Synura petersenii sensu lato*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.