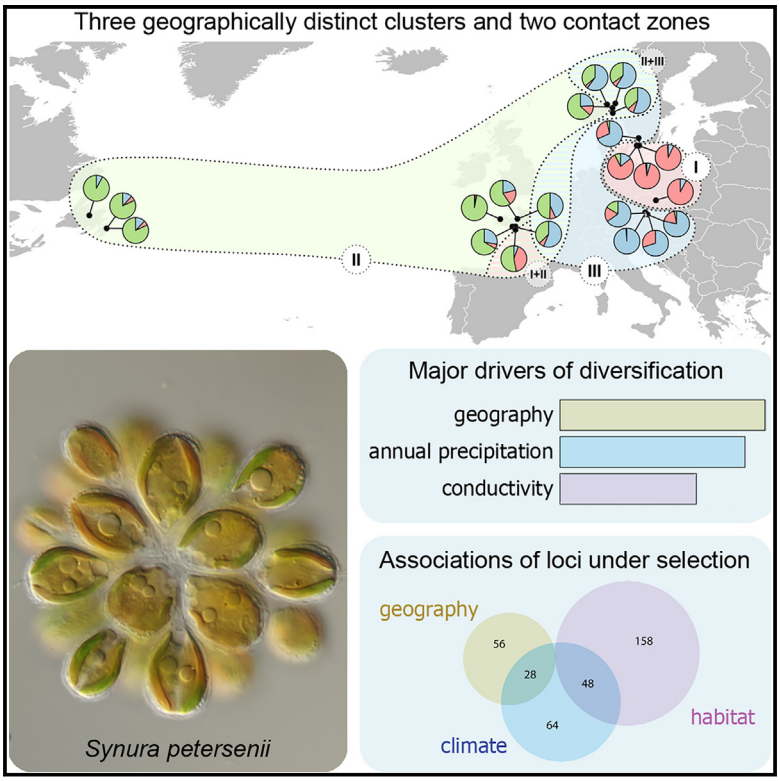
Article

Current Biology

Rapid diversification of a free-living protist is driven by adaptation to climate and habitat

Graphical abstract



Highlights

- *Synura petersenii* has diversified into three population groups with admixture
- Geographical distance drove differentiation at the continental level
- Habitat and climate drove differentiation at the local scale
- Ecological differentiation proceeded despite high dispersal capacity

Authors

Pavel Škaloud, Iva Jadrná, Petr Dvořák, ..., Dora Čertnerová, Helena Bestová, Karin Rengefors

Correspondence

skaloud@natur.cuni.cz (P.Š.), p.dvorak@upol.cz (P.D.)

In brief

Škaloud et al. report the rapid diversification of the photosynthetic protist *Synura petersenii* into three populations. Geographical distance at the continental level, habitat, and climate drove the differentiation of these groups, which are connected with admixture.

Škaloud et al., 2024, Current Biology *34*, 92–105
January 8, 2024
https://doi.org/10.1016/j.cub.2023.11.046

CellPress

CellPress

Current Biology

Article

Rapid diversification of a free-living protist is driven by adaptation to climate and habitat

Pavel Škaloud,[1,6,7,8,*] Iva Jadrná,[1,6] Petr Dvořák,[2,*] Zuzana Škvorová,[1] Martin Pusztai,[1,4] Dora Čertnerová,[1] Helena Bestová,[1,3] and Karin Rengefors[5]

[1]Department of Botany, Faculty of Science, Charles University, 12800 Praha, Czech Republic
[2]Department of Botany, Faculty of Science, Palacký University Olomouc, 78371 Olomouc, Czech Republic
[3]Biodiversity, Macroecology and Biogeography, University of Göttingen, 37077 Göttingen, Germany
[4]Institute for Nanomaterials, Advanced Technologies and Innovation, Technical University of Liberec, 46117 Liberec, Czech Republic
[5]Department of Biology, Lund University, 22362 Lund, Sweden
[6]These authors contributed equally
[7]X (formerly Twitter): @algomicro
[8]Lead contact
*Correspondence: skaloud@natur.cuni.cz (P.Š.), p.dvorak@upol.cz (P.D.)
https://doi.org/10.1016/j.cub.2023.11.046

SUMMARY

Microbial eukaryotes (protists) have major functional roles in aquatic ecosystems, including the biogeochemical cycling of elements as well as occupying various roles in the food web. Despite their importance for ecosystem function, the factors that drive diversification in protists are not known. Here, we aimed to identify the factors that drive differentiation and, subsequently, speciation in a free-living protist, *Synura petersenii* (Chrysophyceae). We sampled five different geographic areas and utilized population genomics and quantitative trait analyses. Habitat and climate were the major drivers of diversification on the local geographical scale, while geography played a role over longer distances. In addition to conductivity and temperature, precipitation was one of the most important environmental drivers of differentiation. Our results imply that flushing episodes (floods) drive microalgal adaptation to different niches, highlighting the potential for rapid diversification in protists.

INTRODUCTION

Eukaryotes are estimated to consist of over 8.7 million species,[1] with the vast majority of the diversity found among the microbial unicellular members, i.e., the protists.[2] Protists are found in virtually all environments on Earth and occupy multiple roles in the food web.[3] Although many are free-living and often photosynthetic, others form various symbiotic associations with other protists and cyanobacteria[4] and act as parasites and pathogens.[5] Photosynthetic protists are important producers of oxygen, form the base of the food web in various ecosystems, and are implicated in many energy and nutrient fluxes.[6,7] Nonetheless, very little is known about the processes that lead to species generation in protists. This is a remarkable gap, considering their importance in ecosystem functioning as well as their abundance. Knowledge of the mechanisms that drive diversification would help to expand the understanding of protist biogeography and biodiversity.

Most concepts and models of speciation, diversification, and adaptation are based on animal and plant systems with obligate sexual reproduction, but to date it is unknown to what extent these can be generalized to microbial speciation.[8] Protists proliferate mainly by asexual reproduction through mitotic cell division, with rare sexual events.[9] Indeed, sexual reproduction has been observed under laboratory conditions in many species,[10–12] implying that protist species are likely separated by reproductive barriers as conceptualized by Mayr in the biological species concept.[13] However, actual mating studies or hybrid zones are not always possible and, in the majority of protist species, mating cannot be induced in the laboratory. Therefore, speciation processes in protists are generally investigated by detection of among-population genetic differentiation indicating a restriction of gene flow due to hitchhiking of alleles or genetic drift facilitated by the prevalence of asexual reproduction.[14,15] Moreover, the existence of reproductive barriers in protists is challenged by their huge population sizes and potentially unlimited dispersal,[16,17] as these attributes should virtually eliminate the presence of barriers to gene flow and prevent the emergence of population differentiation.[18] Accumulating molecular data, however, provide evidence for high levels of population differentiation in protist species, which show geographical,[19,20] ecological,[21,22] or temporal patterns.[23,24] It is plausible that speciation can occur while the gene flow is still present (e.g., by introgression between the populations and species), as it was recently demonstrated in several animal models.[25,26] Because habitat selection may stimulate speciation, it is additionally advantageous to study local adaptations in quantitative traits.[27] However, due to the scarcity of morphological features in protists, non-phenotypic traits are usually investigated, such as growth responses to various environmental variables.[28] Furthermore, several studies have confirmed the existence of evolutionarily very young protist species,[29,30] indicating that the speciation rates of protists may

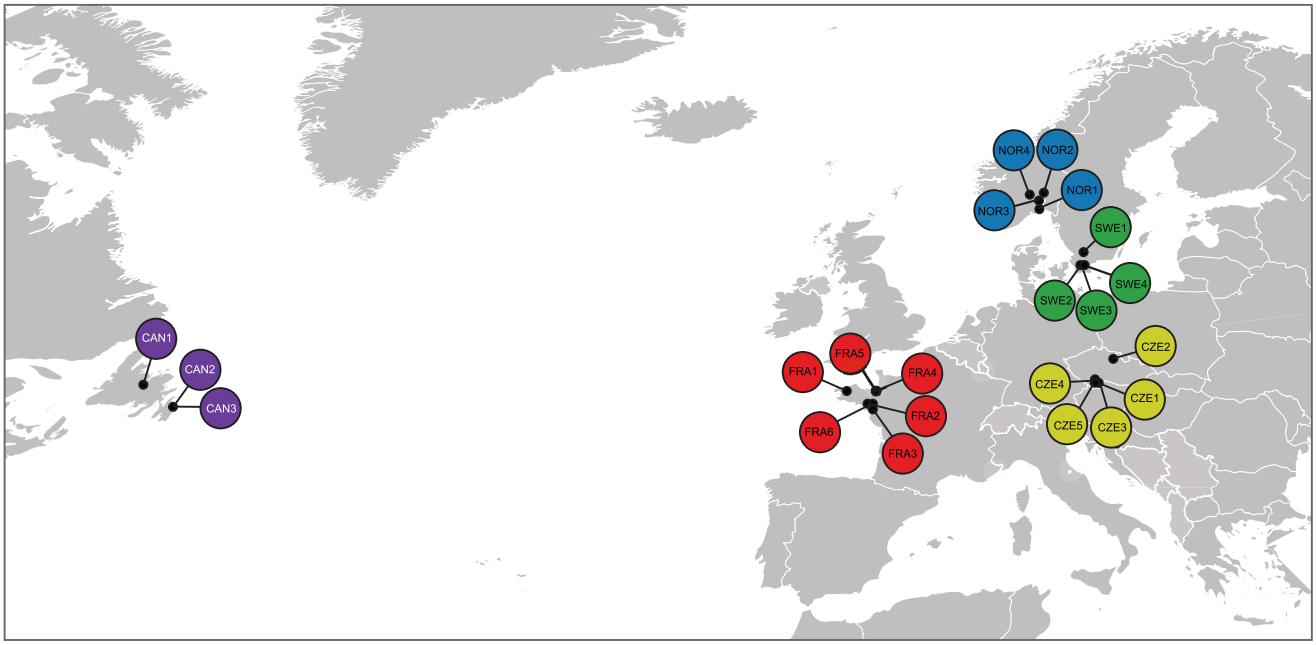


Figure 1. Map of sampled freshwater localities
A total of 22 populations were investigated in 5 distinct geographical regions (color coded).

be equivalent to those of macroorganisms (10,000 to 1 million years).[30] These findings indicate that the assumptions regarding protist dispersal are incorrect and that isolation by physical and/or ecological barriers represents an important driver of protist species divergence.[19]

We investigated *Synura petersenii* (Chrysophyceae, Stramenopiles), a cosmopolitan freshwater golden-brown protist species that is photosynthetic, colonial, and flagellated. This species is particularly well-suited for population differentiation analyses, as it occurs in a great variety of freshwater bodies spanning a broad variety of habitat types.[31] It is also well characterized because numerous studies have focused on its distribution, molecular diversity, morphology, and physiology (e.g., see references in Škaloud et al.[32]). Finally, because of a recognizable fossil record, the origin of this species could be dated to the penultimate glacial period of the mid-Pleistocene.[32] This allows for time estimations of population divergence events.

Synura and other Chrysophyceae often dominate the phytoplankton of meso-oligotrophic temperate habitats[33,34] as they have low demands on temperature, irradiance, and nutrients.[35,36] In general, phytoplankton diversity is maintained by constant changes in abiotic (light, temperature, nutrients) and related biotic factors.[37,38] One of the most prominent factors that changes the structure of freshwater phytoplankton dramatically is the trophic (nutrient level) gradient, which tends to be linked to pH and conductivity.[39,40] Conductivity expresses the concentration of dissolved substances (ions) in water and is usually measured as a proxy value for the trophy or the amount of all dissolved inorganic nutrients. The distribution of *Synura* is considered to be ecologically determined mainly by temperature, pH, and conductivity.[41,42] The solubility and availability of important nutrients, such as phosphorus, change along the pH gradient.[43] Furthermore, the shift of CO_2 forms and their availability in the carbonate buffering system is particularly linked to pH.[44,45] Different phytoplankton species can only use some forms of CO_2[46] and therefore prefer environments with different pH.[44] Chrysophytes lack the carbon concentration mechanism (CCM) that concentrates carbon dioxide around the ribulose bisphosphate carboxylase/oxygenase (RuBisCO) enzyme and, therefore, rely solely on dissolved CO_2 input during photosynthesis.[47,48]

Considering the above, to gain insight into the mechanisms underlying divergence processes in protists, it is essential to identify the factors that affect population structuring on a recent evolutionary timescale. This can only be achieved by combining ecological studies with an assessment of multiple genetic changes in the populations of interest, which would help tease out the impacting factors. Modern genotyping tools, such as restriction-site-associated DNA sequencing (RAD-seq),[49] provide unprecedented opportunities to obtain such insights into the recent evolutionary history of populations in non-model organisms by allowing the investigation of thousands of DNA markers in a population. RAD-seq has recently started to be utilized to investigate the population structure of microalgal protist species,[21,50] providing a much higher resolution of genetic structure than that obtained using other molecular analyses.

Accordingly, in the current study, we aimed to identify drivers of population differentiation in *Synura petersenii* over a short period of time (10,000–150,000 years) using single-digest (sd) RAD-seq. We specifically tested the role of spatial and environmental patterns by investigating the genomic divergence in 22 populations of protists sampled across ecological gradients in 5 geographical regions (Figure 1). In addition, we aimed to uncover whether the recently diverging populations, putatively evolving into incipient species, already show evidence of local adaptation in quantitative traits.

RESULTS

Analysis of organellar loci

Prior to the analysis of RAD-seq data, we used Sanger sequencing of nine mitochondrial and plastid loci to determine the most basal population. Two ecologically most distinct populations were selected in each region. The resulting sequences revealed only minor variation (average p-distance among populations = 0.011) and showed a clear differentiation of the FRA1 population from the other populations (Figure S1). The phylogenetic tree rooted by the closely related *Synura* species clearly identified the FRA1 population as the most basal lineage of all the investigated populations (Figure S2).

RAD-seq and genotyping

sd RAD-seq of 107 *S. petersenii* strains (Table S1) yielded 717 million raw paired-end reads. Of these, 407 million paired-end reads were retained after quality filtering. *De novo* assembly identified 202,056 loci, for which we called 173,776 variant sites. Per-sample depth of coverage ranged from 20× to 666× (mean, 117×). To assess the impact of SNP filtering on the outcome of population genetic analyses, we applied 12 different filtering criteria (described in STAR Methods) for a subset of 10 populations (2 per region). The test datasets differed greatly in the number of loci (562–143,060), SNPs (496–106,183), and percentage of missing data (23%–81%; Table S2). Nonetheless, principal-component analysis (PCA), STRUCTURE, and maximum-likelihood (ML) analyses of these datasets revealed similar clustering and relationships between the strains (Figures S3 and S4). The results were most dissimilar when we analyzed datasets filtered to include loci found in at least one population, most likely because of a high percentage of missing data. In addition, reducing the datasets by including loci found in all but one population resulted in different ordinations of populations in the PCA. Hence, we concluded that at least 50% of populations included should be used as a filtering criterion for downstream population genetic analyses. This filtering strategy, together with writing a single SNP per locus, resulted in keeping 15,792 variant loci (= 9,634 SNPs).

Population genetic diversity and structure

To explore the population genetic structure and genetic diversity in *Synura*, we performed a series of analyses, starting with determining genetic diversity within populations. The basic statistics on population genetic diversity are summarized in Table S3. In general, we detected low levels (mean 2.9%) of polymorphic loci (P_P), suggesting that many SNPs were monomorphic with one allele fixed. We observed the lowest percentage of P_P (1.6%–1.7%) in the FRA1, CAN1, CAN2, and CZE2 populations, indicating that these populations have a lower genetic diversity than the other populations. The FRA3 and FRA6 populations had the highest proportion of private alleles (A_P), which may indicate a certain level of independent evolution of their gene pools. For the variant sites, estimates of observed heterozygosity (H_O) were slightly higher than those of expected heterozygosity (H_E) in all populations (H_O of 0.10–0.23, H_E of 0.06–0.13). However, inbreeding coefficients (F_{IS}) were close to zero (−0.00005 to 0.00008), indicating no apparent deviation from the Hardy-Weinberg equilibrium, suggesting randomly mating populations.

Two different cluster analyses, STRUCTURE and discriminant analysis of principal component (DAPC), predicted the existence of 2–5 clusters of populations, depending on the locus filtering and analysis type (Table S2). The results of these analyses were generally congruent for K = 2 and 3 (where K is the number of populations determined *a priori*), while at higher values of K the analyses showed different clustering of populations (Figure 2). The analyses recovered populations CZE2, SWE2, SWE3, and SWE4 as an independent and well-differentiated cluster I. For K = 3, the remaining populations were separated into clusters II (FRA1, FRA5, FRA6, NOR3, and all Canadian populations) and III (the rest of the populations). The only difference between the methods was in the assignment of the FRA3 population and the presence of historical admixture revealed by the STRUCTURE analysis. Noticeably, at least three strains (E2 in FRA6, P27 in CZE4, and K85 in SWE2 populations) were assigned to a different cluster than the remaining strains of the given population, showing the occasional dispersal of cells between European localities.

PCA (Figure 3A) and principal coordinate analysis (PCoA) plots (Figure 3B) revealed similar clustering patterns of the populations. Both analyses identified a distinct cluster I, consisting of populations CZE2, SWE2, SWE3, and SWE4. The remaining populations formed a large cluster, but with the populations positioned according to the STRUCTURE and DAPC plots. The cluster II and III populations were positioned on the right-hand and left-hand sides of the ordination plots, respectively.

The mutual k-nearest group graphs (mkNNGs) clustering (Figure 4) again recognized the populations CZE2, SWE2, SWE3, and SWE4 as the most distinct cluster (I). The other populations formed a separate network and were grouped into two clusters generally corresponding to the STRUCTURE and DAPC clustering for K = 3. The populations FRA3 and NOR2 were inferred to be members of both clusters II and III.

Phylogeny and geographic structure

We employed Bayesian phylogenetic inference based on RAD-seq data to estimate times for population divergence and to compare the topology with the clustering results described above. The phylogenetic analysis (Figure 5A) inferred the divergence of the *S. petersenii* to 157 (± 21) thousand years ago (kya) and identified several strains unrelated to the remaining strains of the given population. The phylogenetic analysis supported the recognition of the CZE2, SWE2, SWE3, and SWE4 populations as a well-resolved cluster I and identified the cluster II populations FRA1, NOR1, CAN1, CAN2, and CAN3 as the most basal lineage. No other deep lineages were supported by posterior probability values. Importantly, the topology was congruent with the STRUCTURE analysis for K = 3 (Figure 5B) by recognizing clusters II and III, the cluster of admixed populations NOR1, NOR2, NOR4, FRA2, FRA4 (clusters II + III), and the admixed population FRA3 (clusters I + II). Collectively, considering the results of all ordination, cluster, and phylogenetic analyses, we identified a major division of populations into three groups, as well as the presence of several admixed populations (Figure 5C). Cluster I separates four eastern European populations CZE2, SWE2, SWE3, and SWE4, all of them originating from high-conductivity localities (c = 355–892 μS/cm; mean, 527). Cluster II consists of all three Canadian populations and western

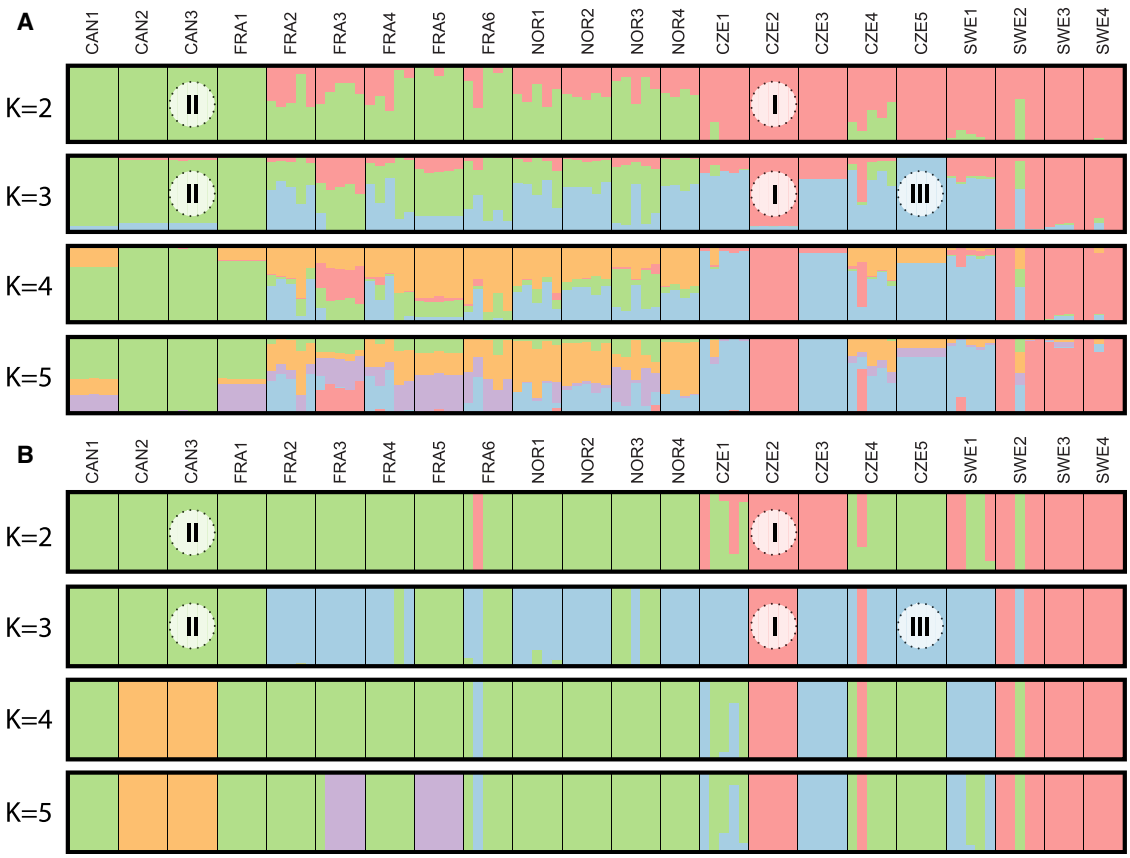


Figure 2. Cluster assignment of the strains

Genetic clustering maps according to STRUCTURE (A) and DAPC (B) analyses, based on SNPs acquired from sdRAD-seq. Each vertical bar corresponds to a strain, with the colors corresponding to the assignment probabilities to each genetic cluster (K). The groups identified by K = 2 and K =3 analyses are indicated by encircled cluster codes (I–III).

See also Figures S1–S3 and Tables S2 and S3.

European populations FRA1, FRA5, FRA6, and NOR3. Cluster III consists of populations CZE1, CZE3, CZE4, CZE5 and SWE1, originating from low conductivity localities (c = 57–185 μS/cm; mean, 103). A total of six populations were identified to be of admixture origin, all originating from localities sampled at the border of clusters II and III. The majority of them (FRA2, FRA4, NOR1, NOR2, NOR4) were recognized to arise by admixture of cluster II and III source populations, whereas the FRA3 population arose by admixture of cluster I and II source populations. Interestingly, this population was sampled at a high-conductivity locality (c = 603 μS/cm).

Genotype-environment associations

Next, we analyzed the association between the genotype and the environment. The outlier detection analysis identified 2,550 loci (=2,550 SNPs) under putative selection, and the gradient forest analysis indicated that geography and climate were the major factors that explained the genetic variation of the loci under selection (Figures 6A and 6B). The first principal coordinates of neighbor matrix (PCNM) axis, representing the spatial structure at the broader spatial scale (Europe vs. Canada), was identified as the most important. Annual precipitation represented the most significant environmental factor. Further, the generalized linear mixed modeling (GLMM) analysis of pairwise population-level genetic distances, with geography, climate, and habitat as the predictor variables, revealed the combined effect of geography and habitat as the major driver of variation (Figure 6C). Habitat was also identified as the major factor by Bayesian hierarchical model (BHM) analysis, which was applied to identify the loci associated with the three predictors above (Figure 6D). We identified 158 loci as correlated with habitat variables only. A significantly smaller number of loci were associated with climate and geography (64 and 56 loci, respectively). In total, 48 loci were associated with both habitat and climate, and 28 loci were associated with both geography and climate.

The PCA of climate-associated loci showed that the loci important for the association were predominantly coupled to annual precipitation, annual mean temperature, and related variables (Figure 6E). The analysis thus suggested that the populations may have diverged as a result of selection connected to precipitation and temperature. The PCA of habitat-associated loci detected conductivity as the major factor behind the population divergence (Figure 6F). Other factors, such as pH, cation exchange capacity, and sand content, may also play a role in habitat selection. Finally, the PCA of loci associated with geography indicated that the Canadian populations may have

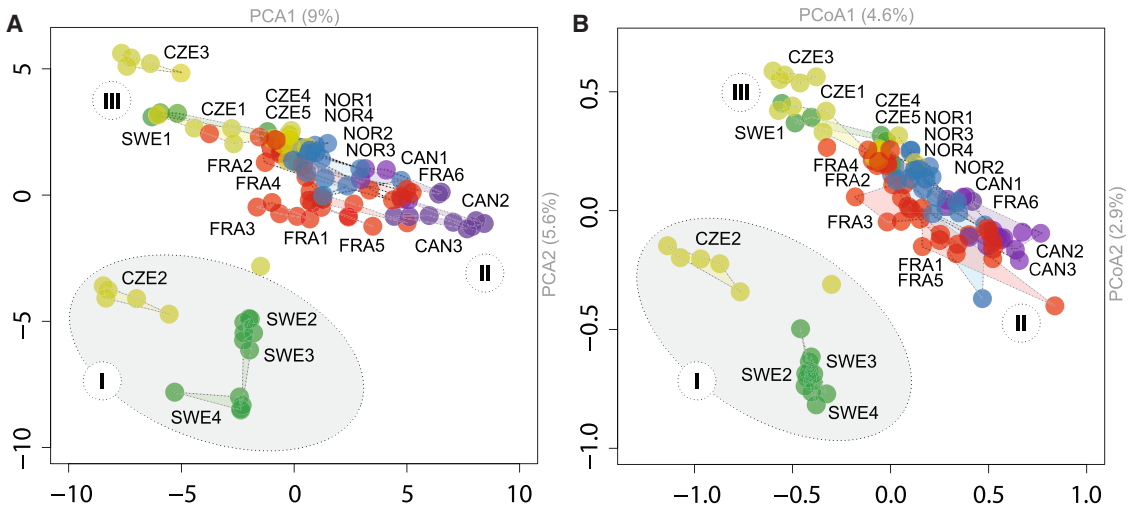


Figure 3. Ordination analyses

PCA (A) and PCoA (B) ordination plots showing genetic relationships between the strains based on SNPs acquired from sdRAD-seq. The strains are color coded based on their geographic origin (see Figure 1) and clustered based on their sampling site, with the exception of obvious immigrant strains. Cluster codes follow the cluster assignment in Figure 2.

See also Figure S3.

diverged from other populations as a result of geographical isolation (Figure 6G).

Population-level differences in quantitative traits

Next, we examined whether selected quantitative traits are differentiated in the identified population clusters. In a subset of ten populations (two per region), we focused on DNA content and maximum growth rate dynamics, as these characteristics are widely thought to play important roles in the ecological strategy of microorganisms.[51–53] We identified substantial variation in DNA content among the populations, ranging from 0.95 to 1.79 Gbp (Table S1). The CAN2 population had the highest DNA content (1.58–1.79 Gbp) and significantly differed among populations (Figure 7A). The DNA content of cluster I populations (CZE2 and SWE2) was slightly larger than that of other related populations. Further, we observed that the responses of population growth rate to temperature were generally congruent among the investigated strains (Figure S5), with the growth optima usually ranging from 15°C to 20°C. However, the growth optimum was significantly suppressed for the CAN1 population, which showed the maximal growth rate at 13.6°C–15.2°C (Figure 7B). Responses of population growth rate to conductivity were much more variable, even between strains from a single population (Figure S6). Nevertheless, we were able to identify some patterns of differentiation among the populations (Figure 7C). SWE1, CZE1, NOR2, and FRA2 exhibited growth optima at intermediate conductivity (395–762 μS/cm). On the other hand, SWE2 strains belonging to cluster I exhibited the optimal growth rate at the highest conductivity tested (2,230 μS/cm). Remarkably, the genetically unrelated F85 (SWE2) strain displayed a preference for lower conductivity. These observations indicated that there were phenotypic differences among populations but that there were also cases of intrapopulation variation and/or migration between populations.

Finally, we performed the variation partitioning analyses to test for the differentiation of clusters in quantitative traits. The tests were performed with either four (I, II, III, and II + III) or six (the basal cluster II divided into three separate lineages) clusters used as a response variable (Figures 7D and 7E). Interestingly, trait differentiation of clusters was much more pronounced when we considered the overall growth dynamics rather than growth optima (Figure 7E). Accordingly, the clusters were significantly differentiated by both, temperature growth dynamics and DNA content, indicating phenotypic differences among the population clusters.

DISCUSSION

Although protists are ecologically important microorganisms, the knowledge of the mechanisms involved in their diversification and speciation is incomplete. This information is critical to understand their distribution patterns, genetic diversity, and what drives differentiation and speciation. We evaluated the factors that could act as drivers of *S. petersenii* differentiation and investigated whether the recently diverged populations, putatively evolving into incipient species, exhibit any local adaptation in quantitative traits. We found that geography, climate, and habitat play an important role in the differentiation of populations. Although geographical isolation has a significant impact on genetic differentiation of distant populations, habitat and climatic factors were the major drivers of population differentiation on a local geographical scale. These findings imply that ecological differentiation allows speciation in microorganisms despite high dispersal capacity and large populations.

Our data suggest the existence of three groups of populations in *S. petersenii* that are in the process of divergence, likely leading to the emergence of new incipient species. Not only do these groups show high genetic divergence but our data also further indicate: (1) restricted gene flow between these groups, low

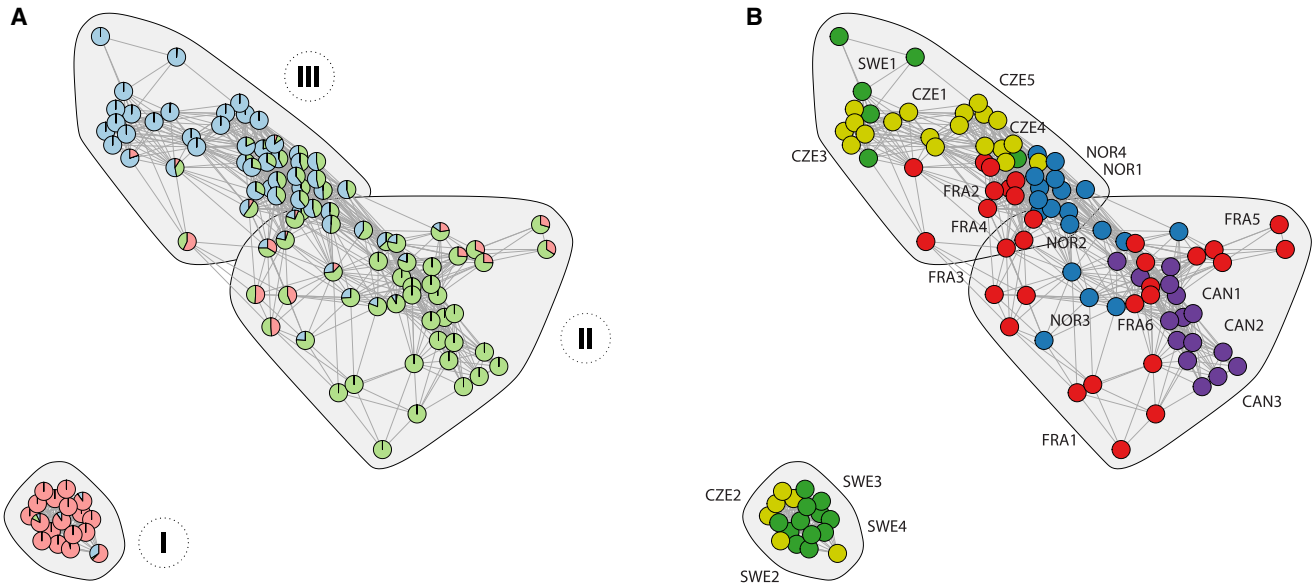


Figure 4. Clustering

The mutual k-nearest group graphs (mkNNGs) of strains, showing the clustering according to the fast greedy algorithm. Edges connect each point to its k nearest points under k = 28. The strains are colored either according to STRUCTURE analysis for K = 3 (A) or based on the geography (B) (see Figure 1). Cluster codes follow the cluster assignment in Figure 2.
See also Figure S3.

admixture inferences, and genome size differences potentially impeding successful mating; and (2) evidence of environmental selection, as indicated by different habitat and climatic preferences and local adaptation to conductivity. Although we cannot compare our results directly with many other microeukaryotes, several well-characterized animal and plant model systems of speciation exhibit similar levels of divergence and local adaptation among their speciating lineages.[54,55] In these cases, the sympatric and allopatric species pairs were reproductively isolated, but with ongoing introgression.

Population genetic structure

We detected high levels of genetic differentiation among populations of the microalga *S. petersenii* and a robust grouping of the investigated strains into several distinct clusters (Figure 5), regardless of the locus filtering strategy used. The FRA1 population from western France represents a clearly separated and most basal lineage. Our clustering analyses (Figures 2, 3, 4, and 5) revealed three major diverged clusters (I–III) with two transitional admixed clusters between them (I + II and II + III). This resembles contact zones; areas where two or more populations potentially hybridize.[56] Contact zones are rarely observed among the microalgae. However, the diatom *Skeletonema marinoi* exhibits two populations adapted to higher and lower salinity, connected by a transitional contact zone along the gradient of salinity between the North and Baltic seas,[56] similar to animals and macroalgae.[57] As in the case of *S. marinoi*, we cannot observe hybrids in nature or in the laboratory, but the genetic and ecological divergence of clusters I–III and transitional admixed clusters suggest a divergence continuum of the populations along the environmental gradients and physical distance. The admixed cluster II + III appears to represent a geographical contact zone situated between the western cluster II and eastern cluster III, and the cluster I + II seems to have arisen by hybridization of the eastern cluster with the distant, high-conductivity cluster I at an ecologically suitable locality. These results illustrate the high dispersal capacities of freshwater microorganisms, probably mediated by passive transport by waterbirds.[58]

The whole Pleistocene period was characterized by climatic oscillations, with repeated glaciations over Europe and North America. The glaciations led to changes in distribution and both effective population size and actual size in plant species.[59] Because we observed a significant genotype-environment association (see below), we assume that the *S. petersenii* population are sensitive to climatic fluctuations. The whole species *S. petersenii* diverged prior to the last glacial period (from 125 to 14.5 kya) and all major and transitional clusters diverged during the glacial period. Interestingly, 15 populations (FRA1, CAN1, CAN2, CAN3, FRA5, NOR2, NOR4, FRA4, CZE3, CZE5, SWE1, FRA3, SWE4, SWE2, and CZE2) radiated after the glacial period ending 14.5 kya. This suggests that the diversification of *Synura* was affected by the quaternary climatic oscillations as Hewitt[60] observed in many plant species (reviewed in Kadereit and Abbott[61]).

Environmental drivers of the *Synura* diversification

The genotype-environment association analysis and RAD-seq clustering patterns observed herein suggest the presence of incipient, mostly environmentally driven diversification in the investigated populations. Climate and habitat factors were identified as the major drivers of differentiation at the local scale, while at larger distances allopatric differentiation seemingly took place as a result of geographical barriers. Generally, very few studies have focused on the factors promoting speciation in free-living

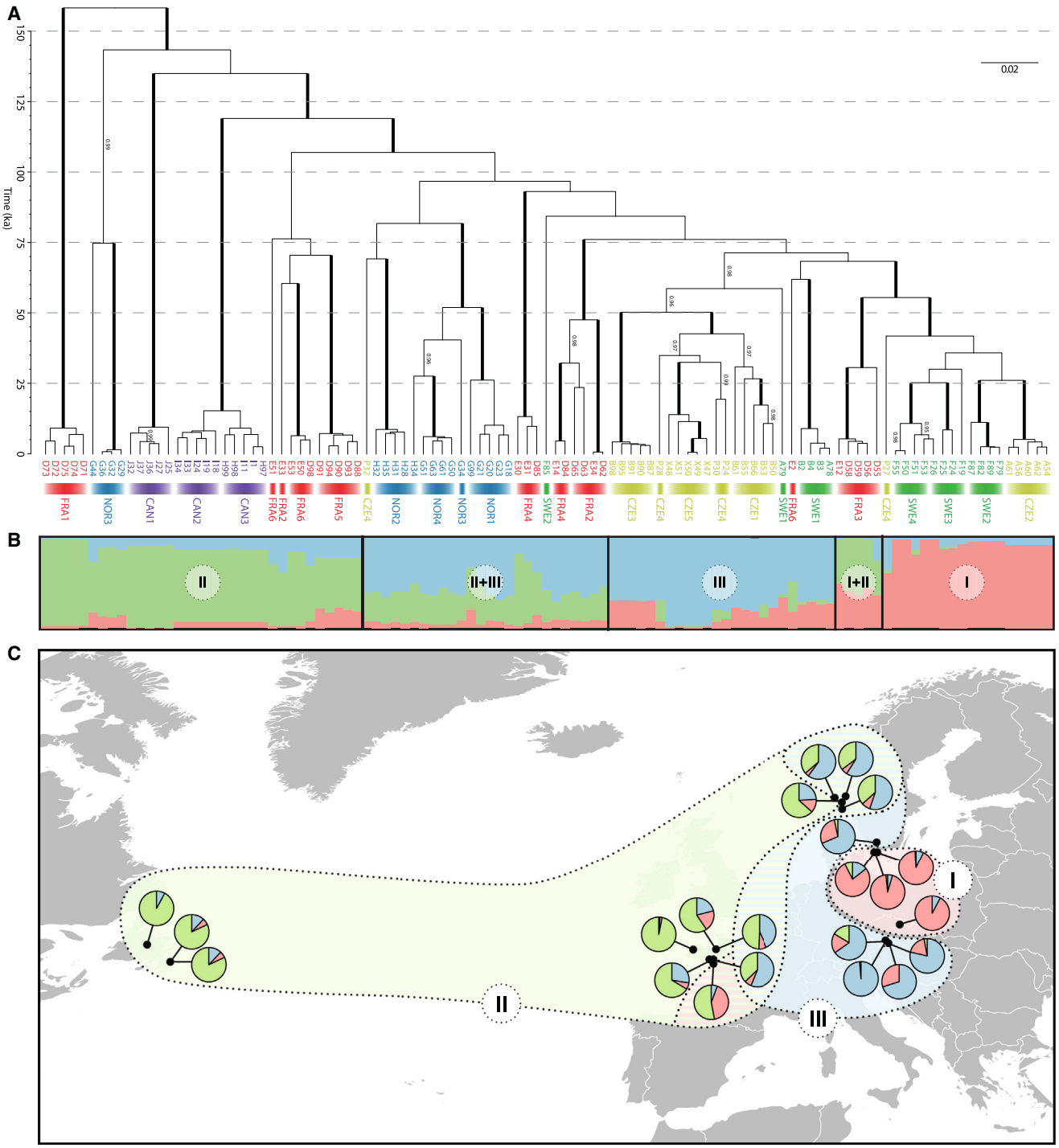


Figure 5. Phylogeny and geographic structure

(A) The BEAST phylogeny based on the analysis of nuclear SNP data, with the FRA1 population specified as an outgroup following the result of organellar loci analysis (see Figure S2). The numbers next to the branches show Bayesian posterior probability (PP) values. The fully supported branches (PP values = 1.00) are thickened. Scale bar represents the expected number of substitutions per site. Time axis is in thousand years ago (kya).

(B) Results of the STRUCTURE analysis for K = 3 mapped in the phylogenetic context.

(C) Geographic distribution of ancestral groups (K = 3). Pie charts show ancestral group proportions, averaged for each locality. Grouping of populations into three population clusters and two admixed clusters is indicated, based on the results of ordination, cluster, and phylogenetic analyses.

See also Figure S4.

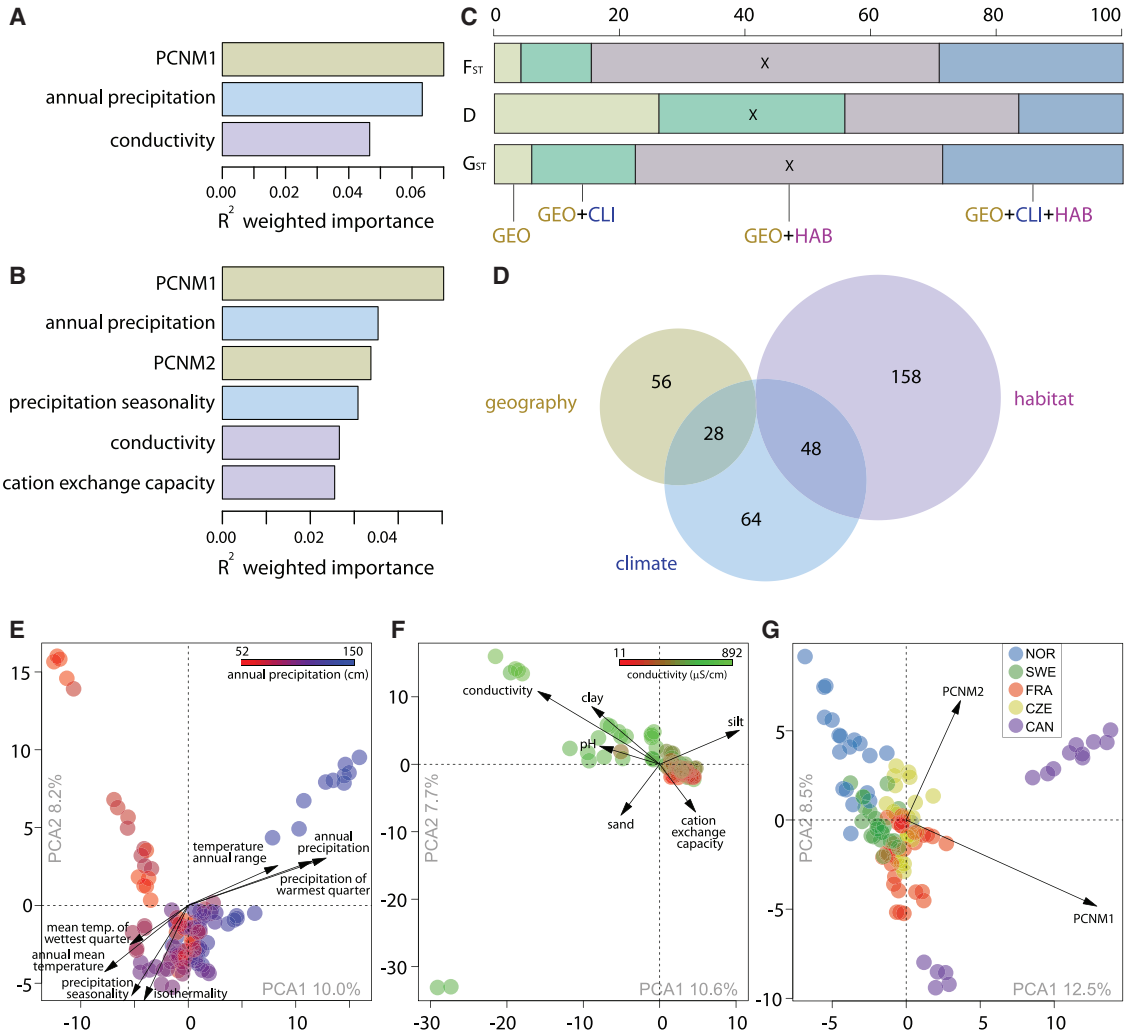


Figure 6. Genotype-environment associations

(A and B) The importance of climate (blue), habitat (violet), and geography (beige) in explaining genetic variation across populations, as determined by the gradient forest analysis of three (A) or six (B) best explaining variables.

(C) Relative contribution of predictors (GEO, geography; CLI, climate; HAB, habitat), based on weighted deviance information criterion (DIC) coefficients from GLMM analysis, explaining differentiation of populations using F_{ST}, G_{ST}, and D_{Jost} genetic divergence metrices. The best-supported models are indicated by an "X."

(D) The number of loci identified by Bayesian hierarchical model (BHM) analysis as significantly associated with climate, habitat, and geographical distance. The number of loci associating with multiple factors are given in intersections between circles.

(E–G) PCA ordination plots showing the associations of loci identified by BHM analysis with selected climatic (E), habitat (F), and geographical (G) predictors. The strains are color coded based on the selected factors characterizing the sampling sites (annual precipitation, conductivity, and geographical area).

protists.[62] These studies have usually investigated the distribution patterns of closely related species or species-level lineages and suggest that both allopatric and sympatric speciation may occur in protists.[30] A few surveys, focusing primarily on foraminifera and diatoms, clearly identified dispersal limitation by geographic distance as likely promoting allopatric speciation processes.[19,27,63] However, a much larger number of studies indicate ecological partitioning of various protist species (including green algae, dinoflagellates, foraminifera, radiolaria, and chrysophytes), implying their origin by sympatric speciation, i.e., without any physical barrier preventing the individuals from mating. In these studies, several ecological factors are suggested as the potential drivers of protist speciation, including salinity,[29,64–66] nutrients,[67,68] light,[69,70] or pH.[28,71,72] Our results corroborate the prevailing view of sympatric speciation by ecological divergence, particularly within a smaller geographical area. However, we focused on much more recent evolutionary processes than those considered in other studies. Considering the genetic and environmental differentiation of cluster I dated to approximately 42 (32–53) kya, we demonstrate that the ecological diversification processes may be very rapid in comparison with the divergence time in other *Synura* species (*S. sphagnicola* 15–17 Ma[30]), as well as other protists such as diatoms (2.9–11.7 Ma[73]), dinoflagellates (5.5–32 Ma[74]), and coccolithophores (0.3–5.6 Ma[75]).

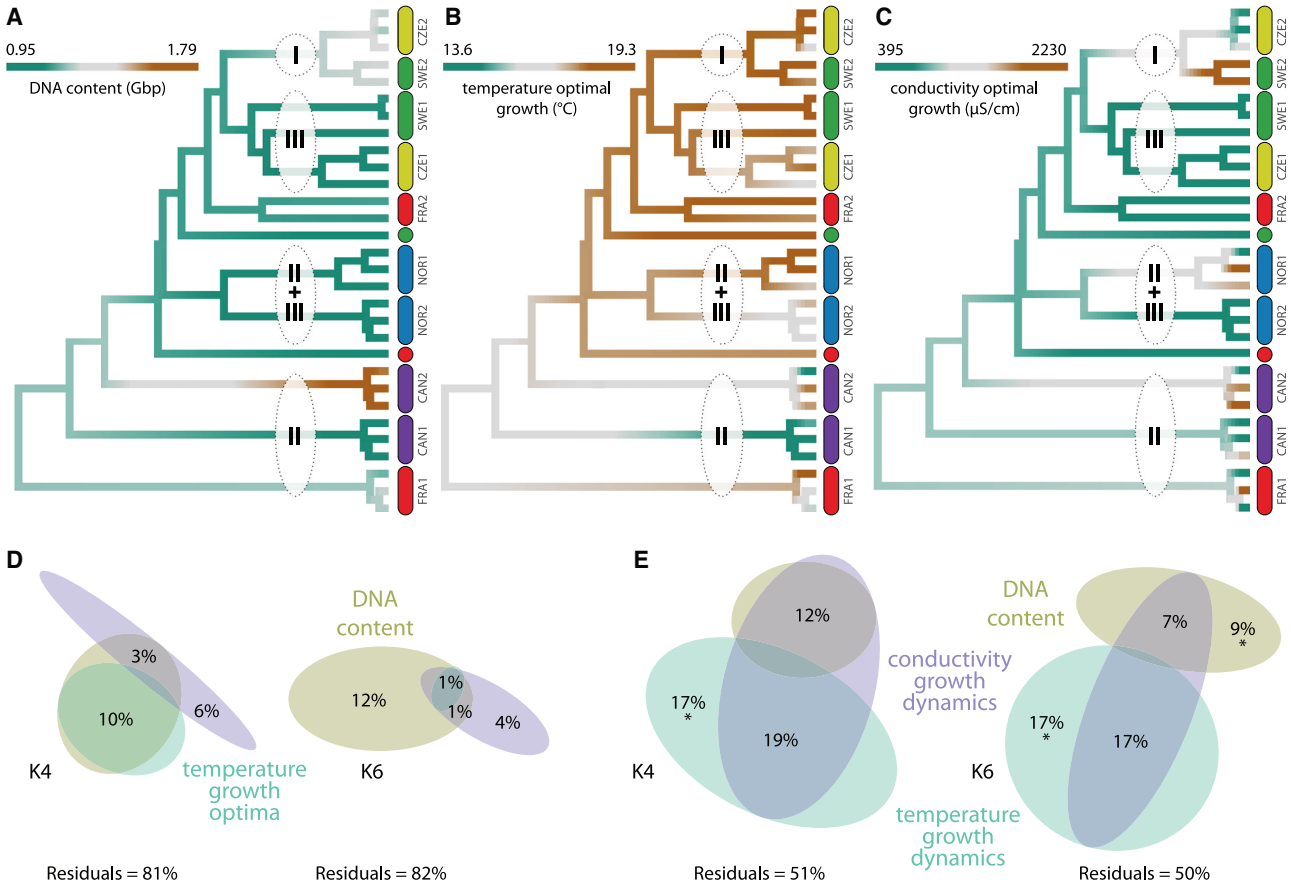


Figure 7. Analysis of quantitative traits

(A–C) Estimated evolution and differences among the ten selected populations in terms of cellular DNA content (A), temperature growth optima (B), and conductivity (ionic content) growth optima (C). The strains are color coded based on their population affiliation (see Figure 1) and their grouping into the four clusters is indicated.

(D) Venn diagrams showing the relative effects of DNA content, temperature growth optima, and conductivity growth optima on the differentiation of four (left) or six (right) clusters of populations, as determined by cluster analyses.

(E) Venn diagrams showing the relative effects of DNA content, temperature growth dynamics, and conductivity growth dynamics on differentiation of four (left) or six (right) clusters of populations. Significant net effects (D and E) are indicated by asterisks (*p = 0.05).

See also Figures S5 and S6 and Table S1.

We identified precipitation, temperature, and conductivity as some of the most important environmental variables that may have impacted *S. petersenii* diversification in Europe (Figures 6E and 6F). Both temperature and conductivity represent the most important abiotic factors in aquatic environments as they vary greatly on spatial and temporal scales.[76] Further, temperature is often identified as the most important abiotic factor that directly influences species performance and, indirectly, the strength of species interactions, including competition and predation.[77] Similarly, conductivity is often identified as the major driver of changes in the composition of protist freshwater communities.[78] Further, conductivity as a climate-sensitive variable also responds rapidly to changes in temperature and precipitation.[79] Indeed, both these factors are repeatedly identified as the key variables controlling the assemblage of freshwater phytoplankton,[80,81] including chrysophytes such as *S. sphagnicola*,[30] a species related to *S. petersenii*.[82,83] Interestingly, precipitation is generally not identified as an important factor that influences protist communities. However, flushing episodes (floods after heavy rains) may greatly impact phytoplankton communities by upwelling nutrients, decreasing transparency, affecting vertical stratification, and removing grazers.[76] Furthermore, climatic changes affecting the balance of precipitation during the year can have a major impact on the transport of nitrogen and phosphorus into rivers and the subsequent eutrophication of lakes.[84,85] The rainy season can cause a rapid and significant increase in nutrient content and subsequent change in the composition and structure of the phytoplankton of the lentic ecosystem.[86,87] Local adaptation to any of the above-mentioned environmental variables may be related to mechanisms that effectively prevent immigrant genotypes from being incorporated into local populations, such as the priority effect,[88] resulting in rapid sympatric speciation. In addition, there is growing evidence that species can diverge even in the presence of gene flow.[25,89] Accordingly, we hypothesize that while some European

populations have differentiated by adaptation to low temperature and factors related to high precipitation levels, the recent diversification of cluster I was promoted by local adaptation to environments with high conductivity, which often relates to high nutrient levels.

Evidence of ecological diversification and adaptation

The extent of local adaptation is critical for the speciation processes.[90] A combination of genetic differences and quantitative trait variation can be used as indirect evidence for the role of selection and may help to identify patterns of local adaptation.[91] Considering the scarcity of morphological features in protists, morphological trait analyses generally cannot be used to detect the patterns of local adaptation. Only very few reports have been published on the adaptive morphological differentiation in protists, including a recent divergence of dinoflagellates after a marine-freshwater postglacial transition[29,66] or selection toward larger coccolith size in the haptophyte genus *Gephyrocapsa*.[92] Consequently, we focused on other phenotypic traits, namely, intrinsic differences in fitness-related ecophysiology and DNA content, to identify phenotypic differences in the populations in the current study.

Although we cannot infer the ancestral ecological conditions of the *S. petersenii* populations based on fossil records, we show that temperature growth dynamics are the most important factor explaining the differences between diverged *S. petersenii* populations (Figure 7E). The basal species clusters represent temperature generalists, whereas the most recent species are temperature specialists, requiring a relatively narrow range of temperature, trending toward warm conditions. Such narrowing of the fundamental niche is usually caused by negative biotic interactions, such as competition and predation.[93] Accordingly, the observed differences in temperature growth dynamics may reflect local adaptations to biotic interactions, recently shown to play a major role in shaping protist community patterns.[94] Moreover, we observed within-cluster variability of the conductivity optima (Figures 7C and S6). This is most liked explained by colonization of a cell originally belonging to another population rather than due to standing variation in the population.

In the current study, the genetic clusters identified by genome-wide RAD-seq were also well differentiated by DNA content. In particular, we detected a significant increase in DNA content in all strains belonging to the CAN2 population (Figure 7A). Considering the nearly 2-fold difference in DNA content between CAN2 and other populations, ancient polyploidization followed by genome downsizing is a likely explanation for this difference. Alternatively, these strains could represent transitional diploid stages because an alternation of two ploidy levels was recently revealed in the life cycle of *Synura* species.[95] However, the latter is very unlikely as all three CAN2 strains would have to be in the diploid phase at the time of DNA content estimation. In fact, polyploidization appears to be an important yet barely studied isolating barrier in many protists. To date, it has been detected in green algae,[96,97] diatoms,[98] dinoflagellates,[99,100] and ciliates.[101] In addition, we detected minor differences in DNA content that distinguish cluster I from the other populations and clearly do not reflect the changes in ploidy. Thus, these differences could be a result of transportable element activity, which is considered largely responsible for genome expansion.[102] However, we cannot provide further evidence for the mechanism of the genome size fluctuations without a high-quality reference genome.

It is possible that the observed minor variation in DNA content may mirror a local adaptation of the populations, which is suggested by the significant correlation between the DNA content and cell size.[31] Indeed, cell size is a very important adaptive parameter that impacts numerous aspects of protist ecology, including light absorption, metabolic activity, grazing susceptibility, and sinking rate.[98] However, it should be noted that DNA content and/or cell size may also be a non-adaptive trait and therefore does not necessarily provide evidence of local adaptation.

The current study might have some limitations, largely due to the study system in question. First, we had a relatively small sample size of 22 localities and 107 individuals, which may limit the statistical power of the genotype-environment associations. However, it is worth mentioning that *Synura* species, like many protists, are characterized by their cryptic morphological diversity (e.g., Škaloud et al.[30]), which prevents genotyping tens of strains isolated from a single locality. Because individuals cannot be distinguished based on morphological features before strain isolation, they were selected by Sanger sequencing of the internal transcribed spacer (ITS) rDNA after cultures had been established. Second, the necessity to investigate isolated and cultured strains might limit the observed genetic diversity due to culturing bias. However, the effect of culturing bias in protists is currently not known. Finally, the selected set of quantitative traits may not represent all the major phenotypic divergences caused by adaptive morphological differentiation of uncovered incipient species. Nonetheless, a significant genotype-environment association resolved by several diverse methodological approaches clearly supports our major findings of rapid ecological speciation.

In conclusion, we show that the diversification process in the protist *Synura* is characterized by high genetic differentiation between the evolving lineages, low gene flow, and strong selection by the environment, suggesting that ecological speciation could be an important mode of speciation in protists over smaller geographical distances. These ecological factors include precipitation, temperature, and conductivity, indicating that climatic changes may further drive the diversification in protists. In fact, we conclude that the diversification process can be rapid, as we observed several recent divergence events in the past 150,000 years. Accordingly, the vast biodiversity of protists and other microorganisms can be explained by rapid ecological speciation despite their widespread dispersal.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS

- Organellar loci analysis
- Preparation of single-digest RAD-seq libraries
- DNA content estimation
- Growth rate experiments

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Processing of RAD-seq data
  - Patterns of diversity
  - Population structure analyses
  - Phylogenetic analyses
  - Detection of loci under putative selection
  - Genotype-environment association analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.11.046.

ACKNOWLEDGMENTS

The authors are grateful to Lenka Flašková, Gabriela Fuxová, Štěpánka Hrdá, Blanka Hamplová, Marian Novotný, Martin Převorovský, and Jakub Vlček for their advice, suggestions, and valuable discussions concerning the preparation of the RAD-seq libraries. We thank Zuzana Münzbergová and Martin Weiser for their advice on statistical analyses and Jens Boenigk, Hanna Johannesson, and Jan Kollár for valuable discussions on the speciation of protists. We would also like to thank Magda Škaloudová and Anna Hirnerová for their help with sampling and growth rate experiments and Marie Svensson for initial optimization of the sdRAD-seq protocol. This research was supported by the Czech Science Foundation (grant no. 23-06881S to P.Š.), Charles University (project GA UK no. 1498119 to I.J.), and the Swedish Research Council (2012-10-24 to K.R.). Computational resources were provided by the e-INFRA CZ project (ID: 90254) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

AUTHOR CONTRIBUTIONS

P.Š. and I.J. conceptualized the study. I.J., Z.Š., P.D., P.Š., and K.R. designed the methods. P.Š., P.D., and I.J. performed formal analysis. I.J., Z.Š., D.Č., P.Š., and H.B. performed the investigation. M.P., P.Š., D.Č., and I.J. acquired resources. P.Š. and I.J. wrote the original draft. P.Š., P.D., K.R., M.P., I.J., and D.Č. reviewed and edited the paper. P.Š., K.R., and I.J. acquired funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 12, 2023
Revised: July 27, 2023
Accepted: November 22, 2023
Published: December 15, 2023

REFERENCES

1. Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G., and Worm, B. (2011). How many species are there on Earth and in the ocean? PLoS Biol. *9*, e1001127.
2. Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The new tree of Eukaryotes. Trends Ecol. Evol. *35*, 43–55.
3. Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., and Thomas, W.K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol. Evol. *27*, 233–243.
4. Ahmadjian, V., and Parecer, S. (2000). Symbiosis: An Introduction to Biological Associations, Second Edition (Oxford University Press).
5. Caron, D.A., Worden, A.Z., Countway, P.D., Demir, E., and Heidelberg, K.B. (2009). Protists are microbes too: a perspective. ISME J. *3*, 4–12.
6. Sorokin, Y.I. (1999). Aquatic Microbial Ecology — A Textbook for Students in Environmental Sciences (Backhuys Publishers).
7. Worden, A.Z., Nolan, J.K., and Palenik, B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. Limnol. Oceanogr. *49*, 168–179.
8. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates).
9. Rengefors, K., Kremp, A., Reusch, T.B.H., and Wood, A.M. (2017). Genetic diversity and evolution in eukaryotic phytoplankton: revelations from population genetic studies. J. Plankton Res. *39*, 165–179.
10. Sandgren, C.D., and Flanagin, J. (1986). Heterothallic sexuality and density dependent encystment in the chrysophycean alga Synura petersenii Korsh. J. Phycol. *22*, 206–216.
11. Mann, D.G., Chepurnov, V.A., and Droop, S.J.M. (1999). Sexuality, incompatibility, size variation, and preferential polyandry in natural populations and clones of Sellaphora pupula (Bacillariophyceae). J. Phycol. *35*, 152–170.
12. D'Alelio, D., Amato, A., Luedeking, A., and Montresor, M. (2009). Sexual and vegetative phases in the planktonic diatom Pseudo-Nitzschia multistriata. Harmful Algae *8*, 225–232.
13. Mayr, E. (1942). Systematics and the Origin of Species from the Viewpoint of a Zoologist (Columbia University Press).
14. Vanoverbeke, J., and De Meester, L. (2010). Clonal erosion and genetic drift in cyclical parthenogens - the interplay between neutral and selective processes. J. Evol. Biol. *23*, 997–1012.
15. Vanormelingen, P., Evans, K.M., Mann, D.G., Lance, S., Debeer, A.E., D'Hondt, S., Verstraete, T., De Meester, L., and Vyverman, W. (2015). Genotypic diversity and differentiation among populations of two benthic freshwater diatoms as revealed by microsatellites. Mol. Ecol. *24*, 4433–4448.
16. Fenchel, T. (1993). There are more small than large species? Oikos *68*, 375–378.
17. Finlay, B.J. (2002). Global dispersal of free-living microbial eukaryote species. Science *296*, 1061–1063.
18. Fenchel, T., and Finlay, B.J. (2006). The diversity of microbes: resurgence of the phenotype. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 1965–1973.
19. Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.E., Kotaki, Y., Rhodes, L., Lundholm, N., Sabbe, K., and Vyverman, W. (2010). Limits to gene flow in a cosmopolitan marine planktonic diatom. Proc. Natl. Acad. Sci. USA *107*, 12952–12957.
20. Rengefors, K., Logares, R., and Laybourn-Parry, J. (2012). Polar lakes may act as ecological islands to aquatic protists. Mol. Ecol. *21*, 3200–3209.
21. Postel, U., Glemser, B., Salazar Alekseyeva, K., Eggers, S.L., Groth, M., Glöckner, G., John, U., Mock, T., Klemm, K., Valentin, K., et al. (2020). Adaptive divergence across Southern Ocean gradients in the pelagic diatom Fragilariopsis kerguelensis. Mol. Ecol. *29*, 4913–4924.
22. Reich, H.G., Kitchen, S.A., Stankiewicz, K.H., Devlin-Durante, M., Fogarty, N.D., and Baums, I.B. (2021). Genomic variation of an endosymbiotic dinoflagellate (Symbiodinium 'fitti') among closely related coral hosts. Mol. Ecol. *30*, 3500–3514.
23. Rynearson, T.A., Newton, J.A., and Armbrust, E.V. (2006). Spring bloom development, genetic variation, and population succession in the planktonic diatom *Ditylum brightwellii*. Limnol. Oceanogr. *51*, 1249–1261.
24. Lebret, K., Kritzberg, E.S., Figueroa, R., and Rengefors, K. (2012). Genetic diversity within and genetic differentiation between blooms of a microalgal species. Environ. Microbiol. *14*, 2395–2404.
25. Feder, J.L., Egan, S.P., and Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends Genet. *28*, 342–350.
26. Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C.D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. *23*, 1817–1828.

27. Darling, K.F., Kucera, M., Pudsey, C.J., and Wade, C.M. (2004). Molecular evidence links cryptic diversification in polar planktonic protists to Quaternary climate dynamics. Proc. Natl. Acad. Sci. USA *101*, 7657–7662.

28. Weisse, T., Berendonk, T., Kamjunke, N., Moser, M., Scheffel, U., Stadler, P., and Weithoff, G. (2011). Significant habitat effects influence protist fitness: evidence for local adaptation from acidic mining lakes. Ecosphere *2*, 1–14.

29. Logares, R., Rengefors, K., Kremp, A., Shalchian-Tabrizi, K., Boltovskoy, A., Tengs, T., Shurtleff, A., and Klaveness, D. (2007). Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? Microb. Ecol. *53*, 549–561.

30. Škaloud, P., Škaloudová, M., Doskočilová, P., Kim, J.I., Shin, W., and Dvořák, P. (2019). Speciation in protists: spatial and ecological divergence processes cause rapid species diversification in a freshwater chrysophyte. Mol. Ecol. *28*, 1084–1095.

31. Čertnerová, D., and Škaloud, P. (2020). Substantial intraspecific genome size variation in golden-brown algae and its phenotypic consequences. Ann. Bot. *126*, 1077–1087.

32. Škaloud, P., Škaloudová, M., Jadrná, I., Bestová, H., Pusztai, M., Kapustin, D., and Siver, P.A. (2020). Comparing morphological and molecular estimates of species diversity in the freshwater genus Synura (stramenopiles): a model for understanding diversity of Eukaryotic microorganisms. J. Phycol. *56*, 574–591.

33. Nicholls, K.H. (2012). Chrysophyte blooms in the plankton and neuston of marine and freshwater systems. In Chrysophyte Algae: Ecology, Phylogeny, and Development, C. Sandgren, J. Smol, and J. Kristiansen, eds. (Cambridge University Press), pp. 181–213.

34. Reynolds, C.S. (1984). The Ecology of Freshwater Phytoplankton (Cambridge University Press).

35. Kamjunke, N., Henrichs, T., and Gaedke, U. (2007). Phosphorus gain by bacterivory promotes the mixotrophic flagellate Dinobryon spp. during re-oligotrophication. J. Plankton Res. *29*, 39–46.

36. Watson, S.B., and McCauley, E. (2010). Light and bacteria: substitutable energy sources for chrysophyte blooms? Nova Hedwigia, Beih. *136*, 213–230.

37. Hutchinson, G.E. (1961). The paradox of the plankton. Am. Nat. *95*, 137–145.

38. Sommer, U., Gliwicz, Z.M., Lampert, W., and Duncan, A. (1986). The PEG-model of seasonal succession of planktonic events in fresh waters. Arch. Hydrobiol. *106*, 433–471.

39. Watson, S.B., McCauley, E., and Downing, J.A. (1997). Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol. Oceanogr. *42*, 487–495.

40. Padisák, J., Hajnal, É., Naselli-Flores, L., Dokulil, M.T., Nõges, P., and Zohary, T. (2010). Convergence and divergence in organization of phytoplankton communities under various regimes of physical and biological control. Hydrobiologia *639*, 205–220.

41. Siver, P.A. (2012). The distribution of chrysophytes along environmental gradients: their use as biological indicators. In Chrysophyte Algae: Ecology, Phylogeny, and Development, C. Sandgren, J. Smol, and J. Kristiansen, eds. (Cambridge University Press), pp. 232–268.

42. Kristiansen, J., and Škaloud, P. (2017). Chrysophyta. In Handbook of the Protists, Second Edition, J.M. Archibald, A.G.B. Simpson, and C.H. Slamovits, eds. (Springer International Publishing), pp. 331–366.

43. Lucas, R.E., and Davis, J.F. (1961). Relationships between ph values of organic soils and availabilities of 12 plant nutrients. Soil Sci. *92*, 177–182.

44. Moss, B. (1973). The influence of environmental factors on the distribution of freshwater algae: an experimental study: II. The role of pH and the carbon dioxide-bicarbonate system. J. Ecol. *61*, 157–177.

45. Yan, N.D. (1979). Phytoplankton community of an acidified, heavy metal-contaminated lake near Sudbury, Ontario: 1973-1977. Water Air Soil Pollut. *11*, 43–55.

46. Raven, J.A. (1970). Exogenous inorganic carbon sources in plant photosynthesis. Biol. Rev. *45*, 167–220.

47. Maberly, S.C., Ball, L.A., Raven, J.A., and Sültemeyer, D. (2009). Inorganic carbon acquisition by chrysophytes(1). J. Phycol. *45*, 1052–1061.

48. Wolfe, A.P., and Siver, P.A. (2013). A hypothesis linking chrysophyte microfossils to lake carbon dynamics on ecological and evolutionary time scales. Glob. Planet Change *111*, 189–198.

49. Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One *3*, e3376.

50. Rengefors, K., Gollnisch, R., Sassenhagen, I., Härnström Aloisi, K., Svensson, M., Lebret, K., Čertnerová, D., Cresko, W.A., Bassham, S., and Ahrén, D. (2021). Genome-wide single nucleotide polymorphism markers reveal population structure and dispersal direction of an expanding nuisance algal bloom species. Mol. Ecol. *30*, 912–925.

51. Westoby, M., Nielsen, D.A., Gillings, M.R., Litchman, E., Madin, J.S., Paulsen, I.T., and Tetu, S.G. (2021). Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. Ecol. Evol. *11*, 3956–3976.

52. Banse, K. (1976). Rates of growth, respiration and photosynthesis of unicellular algae as related to cell size – a review. J. Phycol. *12*, 135–140.

53. Raven, J.A., Knight, C.A., and Beardall, J. (2019). Genome and cell size variation across algal taxa. Perspect. Phycol. *6*, 59–80.

54. Han, F., Lamichhaney, S., Grant, B.R., Grant, P.R., Andersson, L., and Webster, M.T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. *27*, 1004–1015.

55. Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J., and Turner, G.F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science *350*, 1493–1498.

56. Johannesson, K., Le Moan, A., Perini, S., and André, C. (2020). A Darwinian laboratory of multiple contact zones. Trends Ecol. Evol. *35*, 1021–1036.

57. Sjöqvist, C., Godhe, A., Jonsson, P.R., Sundqvist, L., and Kremp, A. (2015). Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea-Baltic Sea salinity gradient. Mol. Ecol. *24*, 2871–2885.

58. Figuerola, J., and Green, A.J. (2002). Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. Freshwater Biology *47*, 483–494.

59. Comes, H.P., and Kadereit, J.W. (1998). The effect of quaternary climatic changes on plant distribution and evolution. Trends in Plant Science *3*, 432–438.

60. Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. Nature *405*, 907–913.

61. Kadereit, J.W., and Abbott, R.J. (2021). Plant speciation in the quaternary. Plant Ecology & Diversity *14*, 105–142.

62. Hernández-Hernández, T., Miller, E.C., Román-Palacios, C., and Wiens, J.J. (2021). Speciation across the Tree of Life. Biol. Rev. Camb. Philos. Soc. *96*, 1205–1242.

63. Evans, K.M., Chepurnov, V.A., Sluiman, H.J., Thomas, S.J., Spears, B.M., and Mann, D.G. (2009). Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. Protist *160*, 386–396.

64. Lazarus, D. (1983). Speciation in pelagic protista and its study in the planktonic microfossil record: a review. Paleobiology *9*, 327–340.

65. Stock, A., Edgcomb, V., Orsi, W., Filker, S., Breiner, H.W., Yakimov, M.M., and Stoeck, T. (2013). Evidence for isolated evolution of deep-sea ciliate communities through geological separation and environmental selection. BMC Microbiol. *13*, 150.

66. Annenkova, N.V., Hansen, G., Moestrup, Ø., and Rengefors, K. (2015). Recent radiation in a marine and freshwater dinoflagellate species flock. ISME J. *9*, 1821–1834.

67. Seears, H.A., Darling, K.F., and Wade, C.M. (2012). Ecological partitioning and diversity in tropical planktonic foraminifera. BMC Evol. Biol. *12*, 54.

68. Ishitani, Y., Ujiié, Y., and Takishita, K. (2014). Uncovering sibling species in Radiolaria: evidence for ecological partitioning in a marine planktonic protist. Mol. Phylogenet. Evol. *78*, 215–222.

69. Foulon, E., Not, F., Jalabert, F., Cariou, T., Massana, R., and Simon, N. (2008). Ecological niche partitioning in the picoplanktonic green alga Micromonas pusilla: evidence from environmental surveys using phylogenetic probes. Environ. Microbiol. *10*, 2433–2443.

70. Weiner, A., Aurahs, R., Kurasawa, A., Kitazato, H., and Kucera, M. (2012). Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. Mol. Ecol. *21*, 4063–4073.

71. Ryšánek, D., Holzinger, A., and Škaloud, P. (2016). Influence of substrate and pH on the diversity of the aeroterrestrial alga Klebsormidium (Klebsormidiales, Streptophyta): a potentially important factor for sympatric speciation. Phycologia *55*, 347–358.

72. Škaloud, P., and Rindi, F. (2013). Ecological differentiation of cryptic species within an asexual protist morphospecies: a case study of filamentous green alga Klebsormidium (Streptophyta). J. Eukaryot. Microbiol. *60*, 350–362.

73. Brown, J.W., and Sorhannus, U. (2010). A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. PLoS One *5*, 1–11.

74. Leaw, C.P., Tan, T.H., Lim, H.C., Teng, S.T., Yong, H.L., Smith, K.F., Rhodes, L., Wolf, M., Holland, W.C., Vandersea, M.W., et al. (2016). New scenario for speciation in the benthic dinoflagellate genus *Coolia* (Dinophyceae). Harmful Algae *55*, 137–149.

75. Sáez, A.G., Probert, I., Geisen, M., Quinn, P., Young, J.R., and Medlin, L.K. (2003). Pseudo-cryptic speciation in coccolithophores. Proc. Natl. Acad. Sci. USA *100*, 7163–7168.

76. Reynolds, C.S. (2006). The Ecology of Phytoplankton (Cambridge University Press).

77. Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M., and Charnov, E.L. (2001). Effects of size and temperature on metabolic rate. Science *293*, 2248–2251.

78. Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., Lange, A., and Jensen, M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. Metabarcoding Metagenom. *2*, e21519.

79. Redmond, L.E. (2018). Alpine limnology of the Rocky Mountains of Canada and the USA in the context of environmental change. Environ. Rev. *26*, 231–238.

80. Sommer, U., Adrian, R., De Senerpont Domis, L., Elser, J.J., Gaedke, U., Ibelings, B., Jeppesen, E., Lürling, M., Molinero, J.C., Mooij, W.M., et al. (2012). Beyond the plankton ecology group (PEG) model: mechanisms driving plankton succession. Annu. Rev. Ecol. Evol. Syst. *43*, 429–448.

81. McMaster, N.L., and Schindler, D.W. (2005). Planktonic and epipelic algal communities and their relationship to physical and chemical variables in alpine ponds in Banff National Park, Canada. Arct. Antarct. Alp. Res. *37*, 337–347.

82. Siver, P.A., and Lott, A.M. (2012). Biogeographic patterns in scaled chrysophytes from the east coast of North America. Freshw. Biol. *57*, 451–466.

83. Kristiansen, J. (2005). Golden Algae: A Biology of Chrysophytes (Koeltz Scientific Books).

84. Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K.M., Andersen, H.E., Lauridsen, T.L., Liboriussen, L., Beklioglu, M., Özen, A., et al. (2009). Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. J. Environ. Qual. *38*, 1930–1941.

85. Jeppesen, E., Kronvang, B., Olesen, J.E., Audet, J., Søndergaard, M., Hoffmann, C.C., Andersen, H.E., Lauridsen, T.L., Liboriussen, L., Larsen, S.E., et al. (2011). Climate change effects on nitrogen loading from cultivated catchments in Europe: implications for nitrogen retention, ecological state of lakes and adaptation. Hydrobiologia *663*, 1–21.

86. Weyhenmeyer, G.A., Willén, E., and Sonesten, L. (2004). Effects of an extreme precipitation event on water chemistry and phytoplankton in the Swedish Lake Mälaren. Boreal Environ. Res. *9*, 409–420.

87. Thompson, P.A., O'Brien, T.D., Paerl, H.W., Peierls, B.L., Harrison, P.J., and Robb, M. (2015). Precipitation as a driver of phytoplankton ecology in coastal waters: a climatic perspective. Estuar. Coast. Shelf Sci. *162*, 119–129.

88. De Meester, L., Gómez, A., Okamura, B., and Schwenk, K. (2002). The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. Acta Oecol. *23*, 121–135.

89. Nosil, P., Feder, J.L., Flaxman, S.M., and Gompert, Z. (2017). Tipping points in the dynamics of speciation. Nat. Ecol. Evol. *1*, 0001.

90. Weisse, T. (2008). Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. Biodivers. Conserv. *17*, 243–259.

91. Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nat. Rev. Genet. *14*, 807–820.

92. Bendif, E.M., Nevado, B., Wong, E.L.Y., Hagino, K., Probert, I., Young, J.R., Rickaby, R.E.M., and Filatov, D.A. (2019). Repeated species radiations in the recent evolution of the key marine phytoplankton lineage Gephyrocapsa. Nat. Commun. *10*, 4234.

93. Hutchinson, G.E. (1978). An Introduction to Population Ecology (Yale University Press).

94. Bock, C., Jensen, M., Forster, D., Marks, S., Nuy, J., Psenner, R., Beisser, D., and Boenigk, J. (2020). Factors shaping community patterns of protists and bacteria on a European scale. Environ. Microbiol. *22*, 2243–2260.

95. Čertnerová, D., Čertner, M., and Škaloud, P. (2022). Alternating nuclear DNA content in chrysophytes provides evidence of their isomorphic haploid-diploid life cycle. Algal Res. *64*, 102707.

96. Poulíčková, A., Mazalová, P., Vašut, R.J., Šarhanová, P., Neustupa, J., and Škaloud, P. (2014). DNA content variation and its significance in the evolution of the genus Micrasterias (Desmidiales, Streptophyta). PLoS One *9*, e86247.

97. Hoshaw, R.W., Wells, C.V., and McCourt, R.M. (1987). A polyploid species complex in Spirogyra maxima (Chlorophyta, Zygnemataceae), a species with large chromosomes 1. J. Phycol. *23*, 267–273.

98. Von Dassow, P., Petersen, T.W., Chepurnov, V.A., and Virginia Armbrust, E. (2008). Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus Thalassiosira (Bacillariophyceae)(1). J. Phycol. *44*, 335–349.

99. Loper, C.L., Steidinger, K.A., and Walker, L.M. (1980). A simple chromosome spread technique for unarmored dinoflagellates and implications of polyploidy in algal cultures. Trans. Am. Microsc. Soc. *99*, 343–346.

100. Holt, J.R., and Pfiester, L.A. (1982). A technique for counting chromosomes of armored dinoflagellates, and chromosome numbers of six freshwater dinoflagellate species. Am. J. Bot. *69*, 1165–1168.

101. Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature *444*, 171–178.

102. Kidwell, M.G. (2002). Transposable elements and the evolution of genome size in eukaryotes. Genetica *115*, 49–63.

103. Boenigk, J., Pfandl, K., and Hansen, P.J. (2006). Exploring strategies for nanoflagellates living in a "wet desert.". Aquat. Microb. Ecol. *44*, 71–83.

104. Otto, F. (1990). DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. Methods Cell Biol. *33*, 105–110.

105. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. *25*, 1965–1978.

106. Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., et al. (2017). SoilGrids250m: global gridded soil information based on machine learning. PLoS One *12*, e0169748.

107. Škaloud, P., Škaloudová, M., Procházková, A., and Němcová, Y. (2014). Morphological delineation and distribution patterns of four newly described species within the *Synura petersenii* species complex (Chrysophyceae, Stramenopiles). Eur. J. Phycol. *49*, 213–229.

108. Kalendar, R., Khassenov, B., Ramankulov, Y., Samuilova, O., and Ivanov, K.I. (2017). FastPCR: an in silico tool for fast primer and probe design and advanced sequence analysis. Genomics *109*, 312–319.

109. Salzburger, W., Ewing, G.B., and Von Haeseler, A. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. Mol. Ecol. *20*, 1952–1963.

110. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

111. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

112. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257.

113. Rochette, N.C., Rivera-Colón, A.G., and Catchen, J.M. (2019). Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. *28*, 4737–4754.

114. Lischer, H.E.L., and Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics *28*, 298–299.

115. R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

116. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.

117. Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. *11*, 94.

118. Steinig, E.J., Neuditschko, M., Khatkar, M.S., Raadsma, H.W., and Zenger, K.R. (2016). Netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. Mol. Ecol. Resour. *16*, 216–227.

119. Besnier, F., and Glover, K.A. (2013). ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. PLoS One *8*, e70651.

120. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. *15*, 1179–1191.

121. Li, Y.L., and Liu, J.X. (2018). StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol. Ecol. Resour. *18*, 176–177.

122. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

123. Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.

124. Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics *201*, 1555–1579.

125. Kynčlová, A., Škaloud, P., and Škaloudová, M. (2010). Unveiling hidden diversity in the Synura petersenii species complex (Synurophyceae, Heterokontophyta). Nova Hedwig Beih. *136*, 283–298.

126. Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., and Cresko, W.A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. *772*, 157–178.

127. Temsch, E.M., Greilhuber, J., and Krisai, R. (2010). Genome size in liverworts. Preslia *82*, 63–80.

128. Doležel, J., and Bartoš, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. Ann. Bot. *95*, 99–110.

129. Sprouffske, K., and Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics *17*, 172.

130. Wickham, H. (2016). ggplot2 Elegant Graphics for Data Analysis (Use R!) (Springer).

131. Rochette, N.C., and Catchen, J.M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. Nat. Protoc. *12*, 2640–2659.

132. Knaus, B.J., and Grünwald, N.J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. *17*, 44–53.

133. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE) (IEEE), pp. 1–8.

134. Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics *24*, 1403–1405.

135. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. *22*, 1–20.

136. Pembleton, L.W., Cogan, N.O.I., and Forster, J.W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. *13*, 946–952.

137. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904.

138. Luu, K., Bazin, E., and Blum, M.G.B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. *17*, 67–77.

139. Storey, J.D., Bass, A.J., Dabney, A., and Robinson, D.. qvalue: Q-value estimation for false discovery rate control. R package version 2.28.0. http://github.com/jdstorey/qvalue.

140. Dray, S., Legendre, P., and Peres-Neto, P.R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. *196*, 483–493.

141. Lefeuvre, P. (2018). BoSSA: a bunch of structure and sequence analysis. https://cran.r-project.org/web/packages/BoSSA/index.html.

142. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2018). vegan: community ecology package. R package version 2.5-2. https://cran.r-project.org/web/packages/vegan/.

143. Ellis, N., Smith, S.J., and Pitcher, C.R. (2012). Gradient forests: calculating importance gradients on physical predictors. Ecology *93*, 156–168.

144. Hadfield, J.D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. *33*, 1–22.

145. Lexer, C., Wüest, R.O., Mangili, S., Heuertz, M., Stölting, K.N., Pearman, P.B., Forest, F., Salamin, N., Zimmermann, N.E., and Bossolini, E. (2014). Genomics of the divergence continuum in an African plant biodiversity hotspot, I: drivers of population divergence in Restio capensis (Restionaceae). Mol. Ecol. *23*, 4373–4386.

146. Kitada, S., Nakamichi, R., and Kishino, H. (2017). The empirical Bayes estimators of fine-scale population structure in high gene flow species. Mol. Ecol. Resour. *17*, 1210–1222.

147. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. J. Stat. Software *25*, 1–18.

148. Wei, T., and Simko, V. (2017). R package 'corrplot': visualization of a correlation matrix. Version 0.84. https://github.com/taiyun/corrplot.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
WC medium	Boenigk et al.[103]	N/A
Otto I buffer	Otto[104]	N/A
Critical commercial assays		
MyTaq DNA polymerase	Bioline	Cat#BIO-21107
Invisorb Spin Plant Mini Kit	Invitek	Cat#1037100300
Qubit dsDNA High Standard Assay Kits	ThermoFischer Scientific	Cat#Q32854
Sbf1-HF restriction enzyme	New England Biolabs	Cat#R3642S
NEB4 buffer	New England Biolabs	Cat#B7004S
NEB2 buffer	New England Biolabs	Cat#B7002S
rATP 100 mM	Promega	Cat#E6011
dATP 100 mM	Fermentas	Cat#R0141
T4 DNA Ligase 2000U/μl	New England Biolabs	Cat#CM0202M
Klenow Fragment 5U/μl	New England Biolabs	Cat#M0212L
MinElute PCR Purification Kit	Qiagen	Cat#28004
Quick Blunting Kit	New England Biolabs	Cat#E1201L
MagJET Magnetic Bead-Based Nucleic Acid Purification kit	ThermoFischer Scientific	Cat#K2828
Q5 High-Fidelity 2X Master Mix	New England Biolabs	Cat#M0492S
Deposited data		
Raw Sequencing Reads	This study	NCBI: PRJNA756637
Sanger sequences	This study	NCBI: MZ853848- MZ853899, MZ935253-MZ935633, OK001379- OK001430
Alignments	This study	https://doi.org/10.17632/pngtj6ymnp.3
R scripts	This study	https://doi.org/10.17632/pngtj6ymnp.3
R scripts	This study	https://github.com/dvorikus/Synura-RADs
geste2baypass.py	N/A	https://github.com/CoBiG2/RAD_Tools
WorldClim v. 2.1. database	Hijmans et al.[105]	http://www.worldclim.com/version2
SoilGrids database	Hengl et al.[106]	soilgrids.org
Experimental models: Organisms/strains		
Synura petersenii; See Table S1 for the list of investigated strains	This study	N/A
Synura americana: S110.B6	This study	S110.B6
Synura americana: M34	This study	M34
Synura americana: J39	Škaloud et al.[32]	J39
Synura borealis: S110.F3	Škaloud et al.[107]	S110.F3
Synura borealis: J57	Škaloud et al.[32]	J57
Oligonucleotides		
See Table S4 for the list of primers	This study	N/A
Software and algorithms		
FastPCR vs. 6	Kalendar et al.[108]	https://primerdigital.com/fastpcr.html
Haplotype Viewer	Salzburger et al.[109]	http://www.cibiv.at/~greg/haploviewer
RaxML v. 8.1.20	Stamatakis[110]	https://cme.h-its.org/exelixis/web/software/raxml/
FastQC v. 0.11.9	Andrews[111]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Kraken2 v. 2.0.8	Wood et al.[112]	https://ccb.jhu.edu/software/kraken2/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Stacks 2.5	Rochette et al.[113]	https://catchenlab.life.illinois.edu/stacks/
PGDSpider v. 2.1.1.5	Lischer et al.[114]	http://www.cmpg.unibe.ch/software/PGDSpider/
R v. 4.0.2	R Core Team[115]	https://www.r-project.org/
STRUCTURE v.2.3.4	Pritchard et al.[116]	https://web.stanford.edu/group/pritchardlab/structure.html
DAPC	Jombart et al.[117]	http://adegenet.r-forge.r-project.org/
NetView	Steinig et al.[118]	https://github.com/esteinig/netview
ParallelStructure v.2.3.4	Besnier and Glover[119]	https://rdrr.io/rforge/ParallelStructure/man/ParallelStructure-package.html
CLUMPAK	Kopelman et al.[120]	http://clumpak.tau.ac.il/
StructureSelector	Li and Liu[121]	https://lmme.ac.cn/StructureSelector/
IQ-TREE v. 1.6.1	Nguyen et al.[122]	http://www.iqtree.org/
BEAST v. 1.10.4	Suchard et al.[123]	https://beast.community/
BayPass v. 2.1	Gautier[124]	https://forgemia.inra.fr/mathieu.gautier/baypass_public

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and code should be directed to and will be fulfilled by the lead contact, Pavel Škaloud (skaloud@natur.cuni.cz).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genetic data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive under the BioProject: PRJNA756637. Sequences of organellar loci have been deposited in the NCBI, GenBank: MZ853848-MZ853899, MZ935253-MZ935633, and OK001379-OK001430. Multiple alignments of organellar loci and ITS rDNA sequences are freely available on Mendeley Data: https://doi.org/10.17632/pngtj6ymnp.3.
- A custom script used to identify and extract outlier loci is available at GitHub: https://github.com/dvorikus/Synura-RADs. All R scripts used to analyze the data are freely available on Mendeley Data: https://doi.org/10.17632/pngtj6ymnp.3.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

22 populations of a colonial flagellate *Synura petersenii* were collected in Europe and east Canada (Table S1). The populations were collected to maximize the geographic area, as well as climatic and habitat variation. In each of five distinct geographical regions (Canada, Czech Republic, France, Norway, Sweden), from three to six lakes were selected to represent a most diverse localities in terms of the conductivity (ionic content) and pH (measured by WTW340i; WTW GmbH, Weilheim, Germany). Several monoclonal cultures were established from each population by isolating single *S. petersenii* colonies by a micropipette. Cells were transferred into 96-well polypropylene plate containing WC liquid medium + TES buffer[103], and after 14 days the strains were transferred into 50ml Erlenmeyer flasks and further cultivated at 15°C, under constant illumination of 40 mmol photons • m^{-2} • s^{-1}. All strains were genotyped by ITS rDNA sequencing as described previously[125] to confirm their identity. The PCR products were purified with MagJET Magnetic Bead-Based Nucleic Acid Purification (ThermoFischer Scientific, Massachusetts, USA) and sequenced in Macrogen Europe (Amsterdam, Netherlands). Finally, 4-5 strains per population (107 strains in total) were selected for subsequent analyses. All selected strains have identical ITS rDNA sequences, except for the CZE2 population differing by a single nucleotide substitution.

METHOD DETAILS

Organellar loci analysis

Nine highly variable organellar regions were sequenced for all selected strains, plus five strains of two outgroup taxa (*S. americana*, *S. borealis*). Four published plastid and mitochondrial *Synura* genomes (GenBank accessions MH795128-30, AF222718) were used for selecting appropriate regions and the primers were designed using the program FastPCR vs. 6.[108] Seven mitochondrial (cox1, cox1-trnY; trnM-atp6; trnW-trnM; rps7-trnI; atp8-nad4L; trnN-rps12) and two plastid (ycf66-petB; trnP-acpP) regions were finally selected for sequencing. The primer details are provided in the Table S4. DNA was extracted as described previously.[107] PCR amplification was performed in a mix consisting of 1 μl of DNA template (not quantified), 6.7 μl H_2O, 2 μl buffer, 0.1 μl of each forward and reverse primers and 0.1 μl MyTaq polymerase, carried out by initial denaturation at 94°C for 4 min; 30 cycles of denaturation at 94°C for 30 s, annealing at 54°C for 45 s and elongation at 72°C for 30 s; with a final extension at 72°C for 10 min. The PCR products were purified and sequenced as described above. The accession numbers of sequenced loci are provided in the Table S5. Aligned loci were concatenated and those positions with either ambiguous bases or deletions in a big majority of sequences were discarded. Haplotype networks were created by Haplotype Viewer.[109] The phylogenetic tree was inferred using RaxML v. 8.1.20[110] with 40 replicates under the GTRGAMMA model, with two partitions corresponding to chloroplast and mitochondrial loci, using the rapid bootstrapping.

Preparation of single-digest RAD-seq libraries

150 ml of densely grown strains were harvested by centrifugation, the supernatant was removed, and pellets were stored at −80°C. Total genomic DNA was extracted from the pellets using Invisorb Spin Plant Mini Kit (Invitek, Hayward, USA). The quality of extracted DNA was checked on the 1% agarose gel stained with ethidium bromide and its concentration was measured using Qubit dsDNA High Standard Assay Kits (ThermoFischer Scientific, Massachusetts, USA). The protocol for single-digest RAD-seq library was derived from methodology published by Etter et al.[126] and Rengefors et al.[50] We started with 1 μg of genomic DNA which was digested by single restriction enzyme Sbf1-HF (New England Biolabs, Massachusetts, USA) at 37°C for 60 min followed by 80°C for 20 min for enzyme inactivation. The reaction mixture consisted of 40 μl of genomic DNA, 5 μl NEB4 buffer, 0.5 μl of Sbf1-HF and 4.5 μl of ddH_2O. The P1 adapters with unique 7-bp and 10-bp barcodes were then ligated to the restricted samples at 22°C for 120 min. The ligation ended by inactivation of the ligase at 65°C for 30 min and slow cool down at 22°C for 30 min. The ligation Master Mix contained 3 μl of 100 nM barcoded P1 adapters (modified Solexa adapters, Table S4), 1 μl of 10x concentrated NEB2 buffer, 0.5 μl of T4 DNA ligase (200.000 U/ml, New England Biolabs), 0.6 μl of 100 mM rATP (Promega), and 5 μl of ddH_2O. The samples with corresponding P2 barcodes were multiplexed and purified using MinElute PCR Purification Kit (Qiagen). Then, the libraries were exposed to short sonication using M220 Focused-ultrasonicator (Covaris) targeting a size of fragments about 500 bp and checked its process on 1% agarose gel stained with ethidium bromide. The genomic DNA was then prepared for ligation of the P2 adapters with 6-bp barcodes. To convert 5′ or 3′ overhangs the Quick Blunting Kit protocol (New England Biolabs) was implemented. The libraries mixed with 2.5 μl of Blunting Buffer, 2.5 μl of 1 mM dNTP Mix and 1 μl of Blunt Enzyme Mix were hold at 22°C for 30 min finished by 70°C for 10 min. This step was directly followed by adding a polyA tail at the 3′ ends of the blunt phosphorylated DNA. The libraries were incubated with 5 μl of NEB2 buffer, 4 μl Klenow Fragment (5U/μl, New England Biolabs) and 1 μl 100mM dATP (Fermentas) at 37°C for 30 min, cooled down slowly at room temperature and purified with MinElute PCR Purification Kit (Qiagen). Afterwards, the ligation of barcoded P2 adapters on the genomic DNA took place with 5 μl of NEB2 buffer, 1 μl of T4 DNA ligase (200.000 U/ml, New England Biolabs), 1 μl of 100 mM rATP (Promega) at 22°C for about 200 min. The libraries were again purified with both MinElute PCR Purification Kit and MagJET Magnetic Bead-Based Nucleic Acid Purification (ThermoFischer Scientific). The final full amplification was performed with 26 μl of DNA template (not quantified), using 28 μl of Q5 High-Fidelity 2X Master Mix and 1 μl of each forward and reverse 10 μM Solexa primers (Table S4). The amplification was performed in a thermal cycler with initial denaturation at 98°C for 30 s; 20 cycles of denaturation at 98°C for 10 s, annealing at 65°C for 60 s and elongation at 72°C for 45 s; and final extension at 72°C for 5 min. The success of amplification was checked on 1% agarose gel stained with ethidium bromide. The DNA segments between 200 and 600 bp were size selected with Pippin Prep (Sage Science). The concentration of final libraries was measured using Qubit dsDNA High Standard Assay Kits (ThermoFischer Scientific) and equally pooled. Sequencing was performed using Illumina HiSeq2500 with TruSeq v4 chemistry (Macrogen, Korea), using 20% PhiX control.

DNA content estimation

A selection of 30 strains (three representatives from 10 selected populations) were used for DNA content estimation, using a propidium iodide flow cytometry (FCM). Each strain was analysed three times on separate days to minimize the effect of random instrumental shift. In the case of low quality measurements (i.e., G_1 coefficient of variation [CV] > 5 %), the strain preparation and analysis was repeated. One ml of well-grown culture was centrifuged in a MiniSpin centrifuge (5 min, 2040 g; Eppendorf) and the superfluous medium was removed by pipetting. Subsequently, 350 μL of ice-cold nuclei isolation buffer Otto I (0.1 m citric acid, 0.5 % Tween-20)[104] was added to the pellet, causing the release of *Synura* nuclei. The resulting suspension was thoroughly mixed and kept on ice. To release nuclei of a plant standard *Solanum pseudocapsicum* (2C = 2.59 pg),[127] a 20-mg piece of fresh leaf tissue was chopped with a razor blade in a plastic Petri dish with 250 μL of ice-cold Otto I buffer. Both suspensions (with *Synura* and standard nuclei) were thoroughly mixed together and filtered through a 42 μm nylon mesh into a special 3.5 mL cuvette for direct use with the flow cytometer. Following a 20-min incubation at room temperature, 50 μg/ml of propidium iodide, 50 μg/ml of RNase IIA and 2 μl/ml

β-mercaptoethanol were added to the sample. The stained sample was immediately analysed using a Partec CyFlow SL cytometer (Partec GmbH, Münster, Germany) equipped with a green solid-state laser (Cobolt Samba, 532 nm, 100 mW). Measurements on each sample were taken for up to 5,000 particles, and the resulting FCM histograms were analysed using FloMax v. 2.4d (Partec, Münster, Germany). The first *Synura* peak on the FCM histogram was identified as G_1 (vegetative cells), the second peak as G_2 (dividing cells). The absolute nuclear DNA amount (C-value) was calculated as sample G_1 peak mean fluorescence / standard G_1 peak mean fluorescence × standard 2C DNA content.[128]

Growth rate experiments

The same selection of 30 strains was subjected to growth rate experiments. The strains were inoculated simultaneously into 50 ml culture flasks with WC liquid medium and kept at 15°C, under constant illumination of 40 mmol photons • m^{-2} • s^{-1}. After 10 days of acclimation, we inoculated six replicates per strain on 96-well plates to a starting concentration of 30 F_0 with the total volume of 250 μl. For temperature gradient experiments, the plates were kept at seven temperatures of 3, 7, 10, 13.5, 18, 21 and 24.5°C using the Labio thermostat (Praha, Czechia). For nutrient concentration experiments, eight different WC media were prepared as follows. First, two stock WC solutions were prepared: the standard medium[103] with conductivity 216 μS/cm and the 10x concentrated medium with conductivity 2230 μS/cm. Then, the stock solutions were diluted with distilled water to obtain eight media of conductivity 20, 40, 100, 200, 400, 800, 1600, and 2230 μS/cm. The strains were inoculated into the media and kept at 15°C. Cell abundances were estimated using the chlorophyll fluorescence yield (F_0) measured by the closed FluorCam FC 800-C (PSI, Drasov, Czechia). Each strain was analysed in six replicates. F_0 was measured every day (1,080 and 1,260 replicates for temperature and nutrient concentration experiments, respectively) till most strains reached a stationary phase. The growth rates were calculated in R, using the package *growthcurver*.[129] After discarding data from non-growing strains, the data were trimmed to include only those days for which none of the measured strains decreased in its abundance. Since the "r" values were sometimes wrongly generated in slowly growing cultures, the growth rates were characterized by the "auc_e" value, which is the empirical area under the curve of the measurements. The final growth rates were calculated by averaging "auc_e" values across replicates, and the growth rate plots were created using the package *ggplot2*.[130] Maximum growth was calculated by fitting the Loess model. In addition, the growth dynamics of each strain were described by fitting 4-degree polynomial regression curves to maximum growth rate data.

We mapped the evolution of DNA content and growth optima in relation to temperature and conductivity onto the rooted, time-calibrated tree inferred by BEAST v. 1.10.4 as described above. The tree was pruned to 30 strains for which the quantitative traits were measured, and ancestral trait reconstructions were inferred using maximum likelihood (ML) in R, using the *contMap* function of the package *phytools*. Variation partitioning analyses were performed using the package *vegan*. The affiliation of strains into either four or six clusters was used as a response variable, following the results of population structure analyses (if six clusters were applied, we considered tree basal cluster II lineages as individual clusters). As explanatory growth rate variables, we used either growth optima (a single variable) or growth dynamics, characterized as the parameters of polynomial function (intercept and four degrees). Significance of the net effect of explanatory variables was tested as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of RAD-seq data

The demultiplexed reads were quality checked in FastQC v. 0.11.9.[111] All samples proceeded to the subsequent analyses. Since some strains were not axenic, we first removed reads which might originate from putative contaminants. The samples were filtered using Kraken2 v. 2.0.8[112] with default settings against a custom database of organisms detected in some strains: bacteria and a protist *Bodo*. *De novo* assembly and single nucleotide polymorphism (SNP) calling were performed in Stacks pipeline v. 2.5.[113] First, the program *process_radtags* was run to demultiplex the individuals and remove low quality data (Q >30 and 150 bp). Then, the three main parameters of pipeline program *denovo_map.pl* -m -M, and -n were selected based on the published protocol.[131] We tested following parameter combinations: m = 3-5, M = 1-6, and n = 1-6. The combination with the highest number of polymorphic loci and present in 80% of the samples was selected as optimal: m = 5, M = 3, and n = 3. The program *denovo_map.pl* was run using the selected parameters on paired-end data, removing PCR duplicates. To assess the impact of SNP filtering on the outcome of population genetic analyses, we generated 12 test datasets using the *populations* program, applying different values for parameters *p* (minimum number of populations a locus must be present in to process a locus = 1, 5, 9) and *r* (minimum percentage of individuals in a population required to process a locus for that population = 0.2, 0.4), writing either all SNP per locus or restrict the datasets to only the first SNP per locus. The datasets were not phased and remained haploid. All datasets were exported in Structure, genepop and vcf formats for downstream analyses. Other formats were prepared with PGDSpider v. 2.1.1.5[114] if not specified otherwise. After evaluating 12 test datasets (see Results), a filtering strategy *p*=5 *r*=0.2, and writing a single SNP per locus, was applied to generate a final dataset. This filtering strategy received the best consensus of all filtering strategies and downstream analyses performed (Figures S3, S4).

Patterns of diversity

Genetic diversity statistics within populations, including number of private alleles (A_P), percentage of polymorphic loci (P_P), observed (H_O) and expected heterozygosity (H_E), nucleotide diversity (π), and inbreeding coefficient (F_{IS}) were estimated using the *populations* program in Stacks, based on nuclear SNP data. Other analyses were run in R v. 4.0.5.[115] Proportion of missing data and mean read depth were calculated using the package *vcfR*.[132]

Population structure analyses

Population genetic structure was inferred using a Bayesian model-based clustering in STRUCTURE v.2.3.4,[116] a sequential k-means clustering using the Discriminant Analysis of Principal Components (DAPC),[117] and a mutual k-nearest neighbour graphs clustering (mkNNGs) in NetView pipeline.[118] Multiple STRUCTURE analyses were run using ParallelStructure v.2.3.4[119] in CIPRES science gateway.[133] Initially, the analyses were run for 12 test datasets (see above), using K 1-8, 2,000 burn-in, 10,000 Markov chain Monte Carlo (MCMC) generations, and 20 iterations for each K. The final dataset was analysed using K 1-14, 50,000 burn-in, 100,000 MCMC generations, running 100 iterations for each K. Replicates were summarized and visualized using the CLUMPAK server,[120] applying the CLUMPP LargeKGreedy algorithm. Estimates of the best K to describe the data were determined by the Evanno method using the StructureSelector server,[121] as well as by manual visual inspection of clustering results. In general, delta K of STRUCTURE runs were quite low (ranging from 0.1 to 3.6), suggesting STRUCTURE itself is unable to find a significant structure pattern. We obtained very low delta K values (0.1-0.6) For K 6-14, therefore these STRUCTURE plots were not further considered. DAPC analyses were performed in R, using the package *adegenet*.[134] 1,000,000 iterations were run for K 2-6. The number of retained PCs was set to 20 to visualize the admixture. Estimates of the best K were determined with the help of discriminant analysis scatterplots. The mkNNGs analysis was performed in R, using the package *netview*.[118] The Euclidean distance was used as an input. To detect the best k value, we produced a plot of the number of detected clusters across k to indicate a stable assembly of the graph (a minimum range for the k parameter). Accordingly, the network was constructed at k = 28, where all three algorithms (Fast-Greedy, Infomap, Walktrap) show a general congruence in a number of resolved clusters. Further, we displayed genetic distances among individual strains using principal component (PCA) and principal coordinate (PCoA) analyses. The calculations were performed in R, using the packages *adegenet* and *ade4*.[135] PCA analyses were run on both test and final datasets, using the genlight object generated by the *vcfR* package as an input. PCoA analysis was run for a final dataset only, based on Nei's distances counted by the package *StAMPP*,[136] and transformed to Euclidean distances using the cailliez function.

Phylogenetic analyses

Initially, unrooted phylogenetic trees were inferred on 12 test datasets (see above), based on the multiple sequence alignments produced by the *Populations* program in Stacks. Maximum-likelihood tree reconstruction was performed in IQ-TREE v. 1.6.1[122] using GTR+I+G substitution model. The tree topology was tested using ultrafast bootstrapping with 2,000 replications implemented within the same software. The final dataset was used to infer the rooted, time-calibrated tree. The Bayesian evolutionary analysis was performed by BEAST v. 1.10.4,[123] using the GTR+I+G substitution model with 4 gamma categories, the lognormal relaxed clock type, and a birth–death diversification process tree prior. The population FRA1 was specified as an outgroup following the results of organellar loci analyses (see above). The root age was constrained to 0.140 Ma (SD 0.01), following the inferred molecular divergence time of *S. petersenii* published by Škaloud et al.[32] Two MCMC analyses were run, each for 100 million generations (burn-in 20 million generations). The convergence diagnostics was performed in Tracer v. 1.6.[137]

Detection of loci under putative selection

The outlier loci, which were likely to be largely affected by selection, were identified and extracted for the further analyses using a custom-made script (https://github.com/dvorikus/Synura-RADs) using R package *pcadapt* 4.3.3[138] and *qvalue* 2.28.0.[139] *Pcadapt* is not sensitive to the hierarchical population structure as other approaches, and it can also handle admixed individuals.[138] The input vcf file was produced by Stacks (see above), without any outlier filtering strategy applied. The number of principal components was set to K = 5 based on scree plot and PCA in *pcadapt* produced by the *pcadapt* package itself. We also tested K = 4 and the extreme value of K = 20 and obtained consistent results in the downstream analyses (data not shown). The outlier loci were identified based on q-value and with p-value corrected by Benjamini-Hochberg procedure. Both approaches showed the same set of the 2,550 outlier loci (= 1.26% of all loci).

Genotype-environment association analyses

Three approaches have been employed to detect the correlation between the allele frequencies of the outlier loci and climatic factors, habitat, and geography. The gradient forest analysis and a generalized linear mixed modelling (GLMM) were employed to assess the relative effects of climate, habitat, and geography on the genetic diversity, and a Bayesian hierarchical modelling (BHM) was applied to identify loci associated with the three above-mentioned factors. The strains were isolated, cultivated and genetically characterized by RAD sequencing as specified above. To avoid analyses of linked SNPs, the single SNP was selected per each outlier locus. Accordingly, a total of 2,550 SNPs were analyzed. Explanatory climatic factors were represented by 19 bioclimatic variables obtained from the WorldClim v. 2.1. database[105] at resolution of 2.5 arc minutes. Habitat factors were represented by measured values of pH and conductivity (ionic content), and by seven physical and chemical soil properties obtained from the SoilGrids database.[106] At every sampling site, climatic and habitat data were obtained by applying a 2-km buffer to limit the effects of spatial bias. Geographical distances were transformed to the principal coordinates of neighbour matrices (PCNM)[140] in R, using the packages *BoSSA*[141] and *vegan*,[142] applying a threshold of 1,100 km to get appropriate PCNM scores. All variables were centred and standardized.

The gradient forest analysis was carried out in R v. 3.6, using the packages *gradientForest* and *extendedForest*.[143] First, variation in allele frequencies of 2,550 outlier SNPs was summarized using PCA as described above. The first 80 PCA axes were used as a response variable. To reduce the number of explanatory variables, two most significant climatic and habitat factors were selected by the forward selection in redundancy analysis (RDA) using the package *vegan*. The most significant variables included annual

precipitation (BIO12), precipitation seasonality (BIO15), conductivity and cation exchange capacity. First and second PCNM axes were used as geographical predictors. Two gradient forest analyses were performed, using either six explanatory variables as described above, or three variables only, representing the most important predictors (annual precipitation, conductivity, PCNM1). To take the hierarchical sampling design into account, the locality was set as a random factor. 300 separate analyses were conducted based on 26 randomly selected samples (one per each locality), and the models were combined using the ''combinedGradientForest'' function. Gradient models were fitted with 500 trees, the correlation threshold set to 0.5 and maxLevel = 3.

GLMM analyses were performed in R v. 4.1.0, using the package $MCMC_{GLMM}$,[144] running the modified script published by Lexer et al.[145] F_{ST}, G_{ST}, and D_{Jost} genetic pairwise distances were used as response variables, based on outlier loci only. F_{ST} was estimated using the *populations* program in Stacks, the other two indices were calculated in R, using the package *FinePop*.[146] As climatic and habitat predictor variables, we used Euclidean distances retrieved from PCA analyses of all 19 bioclimatic and 9 habitat variables, respectively. As geographical predictors, we used geospheric distances between populations. Locality pairs were set as a random factor, using the *idv* variance function. The analyses accounted for nonindependence in the data by using pairwise matrices of population-level metrics. Eight different models resulted from combinations of the three predictor variables: a null model without any predictor, three models with a single predictor, three with different combinations of two predictors and one with all three predictors. The deviance information criterions (DIC) and associated DIC differences and weights were used to compare all models. The analyses were run with a burn-in of 200,000 followed by 2 mil iterations with a thinning interval of 750.

The BHM analyses were run using the program BayPass v. 2.1,[124] using default parameters under the standard covariate model. Population allele counts were used as input data. The input file was generated using the *geste2baypass.py* python script (https://github.com/CoBiG2/RAD_Tools). The BayPass method assumes Hardy-Weinberg equilibrium, which can be violated under predominant asexual reproduction. However, our analyses indicate no apparent deviation from the Hardy–Weinberg equilibrium in most of the populations indicating relatively frequent sexual reproduction, so the loci are not likely linked. As climatic and habitat predictor variables, we used the first axis of a PCA on all 19 bioclimatic and 9 habitat variables, respectively. First PCNM axis was used as geographical predictor. The BayesFactors estimates were calibrated by creating the pseudo-observed data sets with 1,000 SNPs, using the *simulate.baypass* function. SNPs with a false detection rate (q-value) under 0.05 were considered as significantly associated with the predictors, using the R function quantile. Finally, the loci under selection detected by BHM analyses were used as input to the PCA using environmental and spatial variables as supplementary variables to show their associations with PCA axes. The PCA analyses were run in R, using the package *FactoMiner*.[147] Because PCA does not allow any missing data, we imputed missing values using the most common genotype at each SNP across all individuals. Prior the analysis, we reduced the number of supplementary environmental variables by excluding correlated variables (Spearman rank correlation > 0.75). Variables to retain were selected by the inspection of PCA ordination plots and corrplots made in R, using the package *corrplot*.[148] The final set of retained variables included seven climatic and six habitat predictors, as follows: annual mean temperature (BIO1), isothermality (BIO3), temperature annual range (BIO7), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12), precipitation seasonality (BIO15), precipitation of warmest quarter (BIO18), pH, conductivity, cation exchange capacity, sand, clay, and silt contents. The first two PCNM axes were used as geographical predictors.