MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2955–2972 doi: 10.1111/mec.13218

Hidden diversity in the freshwater planktonic diatom *Asterionella formosa*

S. VAN DEN WYNGAERT,*† ‡[1] M. MÖST,*† §[1] R. FREIMANN,¶ B. W. IBELINGS*†** and P. SPAAK*†

*Aquatic Ecology, Eawag, Ueberlandstrasse 133, PO Box 611, CH-8600 Duebendorf, Switzerland, †Institute of Integrative Biology, ETH Zürich, CH-8092 Zürich, Switzerland, ‡Experimental Limnology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB) Berlin, Alte Fischerhütte 2, 16775 Stechlin, Germany, §Department of Zoology, University of Cambridge, Downing St., Cambridge CB2 3EJ, UK, ¶Institute of Molecular Health Sciences, ETH-Zürich, Otto-Stern-Weg 7, CH-8093 Zürich, Switzerland, **Institut F.-A. Forel, Université de Genève, 10 Route de Suisse, CH-12090 Versoix, Switzerland

Abstract

Many freshwater and marine algal species are described as having cosmopolitan distributions. Whether these widely distributed morphologically similar algae also share a similar gene pool remains often unclear. In the context of island biogeography theory, stronger spatial isolation deemed typical of freshwater lakes should restrict gene flow and lead to higher genetic differentiation among lakes. Using nine microsatellite loci, we investigate the genetic diversity of a widely distributed freshwater planktonic diatom, *Asterionella formosa*, across different lakes in Switzerland and the Netherlands. We applied a hierarchical spatial sampling design to determine the geographical scale at which populations are structured. A subset of the isolates was additionally analysed using amplified fragment length polymorphism (AFLP) markers. Our results revealed complex and unexpected population structure in *A. formosa* with evidence for both restricted and moderate to high gene flow at the same time. Different genetic markers (microsatellites and AFLPs) analysed with a variety of multivariate methods consistently revealed that genetic differentiation within lakes was much stronger than among lakes, indicating the presence of cryptic species within *A. formosa*. We conclude that the hidden diversity found in this study is expected to have implications for the further use of *A. formosa* in biogeographical, conservation and ecological studies. Further research using species-level phylogenetic markers is necessary to place the observed differentiation in an evolutionary context of speciation.

Keywords: amplified fragment length polymorphism, *Asterionella formosa*, cryptic diversity, diatom, microsatellites

Received 24 February 2015; revision received 23 April 2015; accepted 24 April 2015

Introduction

Many freshwater and marine algal species are described as having cosmopolitan distributions which seems to be in accordance with Beijerinck's (1919) and Baas Becking's (de Wit & Bouvier 2006) famous metaphor 'in micro-organisms everything is everywhere, but, the environment selects', later emphasized by Fenchel & Finlay (2004). However, several studies contradict this hypothesis and report a distinct biogeographical distribution for microbes (Medlin 2007; Vanormelingen *et al.* 2008; Casteleyn *et al.* 2010). Whether these widely distributed morphologically similar algae also share a common gene pool often remains unclear (Medlin 2007). However, this is an important question as intraspecific diversity (genetic polymorphism) represents the evolutionary and adaptive potential of a species. Restricted gene flow, due to geographical and/or intrinsic or extrinsic reproductive barriers, may lead to the fragmentation of the common gene pool and local

Correspondence: Silke Van den Wyngaert, Fax: +33 082 69917; E-mail: silke@uglylab.eu

[1]These authors contributed equally to this article.

adaptation of populations which may eventually lead to the evolution of new (often cryptic) species. Therefore, knowledge on gene pool size (genetic diversity) and gene flow between pools (population structure) is critical for speculating that everything is everywhere.

Population genetic studies of microalgae using high-resolution genetic markers are still scarce relative to other organisms, and most of those studies so far have dealt with marine species (Rynearson & Armbrust 2000, 2004, 2005; Iglesias-Rodriguez *et al.* 2006; Rynearson *et al.* 2006, 2009; Alpermann *et al.* 2009; Nagai *et al.* 2009; Erdner *et al.* 2011; Härnström *et al.* 2011). A general assertion that marine environments offer few physical barriers to dispersal has led to the assumption that protist populations are genetically homogeneous. Two studies on the marine diatom *Pseudo-nitzschia pungens* are in agreement with the assumption of panmixia and found evidence for a single largely unstructured population within two large but heterogeneous areas of the North Sea (Evans *et al.* 2005; Casteleyn *et al.* 2009). However, the majority of studies have shown opposite results and report both large- and fine-scale spatial and temporal population genetic structure within a variety of marine protist species (Rynearson *et al.* 2006; Alpermann *et al.* 2009; Nagai *et al.* 2009; Lowe *et al.* 2010). Results on fine-scale population differentiation need to be interpreted with care as similar patterns can also be caused by the presence of coexisting cryptic species. This was actually the case in the study of Rynearson *et al.* (2006) where population differentiation was driven by the presence of two cryptic *Ditylum brightwelli* species as later shown by Koester *et al.* (2010). Careful investigation of species limits, using a combination of molecular markers with species-level resolution and/or fine-grained morphological analysis and/or mating experiments (Beszteri *et al.* 2007; Casteleyn *et al.* 2008; Evans *et al.* 2009; Poulickova *et al.* 2010), is necessary as incorrect definitions of evolutionary units may lead to erroneous conclusions in population genetic studies (Pante *et al.* 2015).

Freshwater lakes are considered to represent more discrete, patchy and isolated habitats compared to the open ocean [but see, e.g. Leibold & Norberg (2004) for lake plankton biodiversity in a metacommunity context]. Consequently, it is hypothesized that this stronger spatial isolation restricts gene flow, leading to genetically more distinct protist populations between lakes compared to marine habitats. The handful of studies on genetic differentiation among and within freshwater protist populations seems to agree with this hypothesis. Microsatellite analyses on the benthic freshwater diatom *Sellaphora capitata* found highly differentiated populations between lakes in the UK, Belgium and Australia (Evans *et al.* 2009). Similar results were obtained for the planktonic diatom *Fragilaria capucina*, analysed by random amplified polymorphic DNA (RAPD), showing genetically highly differentiated populations among seven lakes along a latitudinal gradient across North America (Lewis *et al.* 1997). Genetically distinct populations of planktonic freshwater diatoms were also found on smaller spatial scales. Based on amplified fragment length polymorphism (AFLP) data, two distinct populations of *Cyclotella meneghiniana* were detected in the same river, within 40 km distance (Beszteri *et al.* 2007), whereas also AFLP-based genotypes of *Asterionella formosa* populations of two geographically close Dutch lakes clustered according to their lake of origin (De Bruin *et al.* 2004). However, a contrasting example is the low differentiation of a ciliate population in four closely located Chinese lakes (Zhang *et al.* 2006). Eukaryotic microalgae commonly reproduce asexually, alternated with rare sexual events (Weisse 2008). Despite the dominance of clonal reproduction, the majority of studies find surprisingly high levels of genetic diversity using polymorphic markers such as AFLP and microsatellites (Rynearson & Armbrust 2000; Logares *et al.* 2009; Lowe *et al.* 2010). For instance, even during bloom conditions clonal diversity of the marine centric diatom *D. brightwellii* remained high with 87–95% distinct genotypes (Rynearson & Armbrust 2005).

In this study, we examine the genetic diversity and population structure of the freshwater planktonic diatom, *A. formosa*. *Asterionella formosa* is one of the dominant algal species in many freshwater lakes during the phytoplankton spring bloom. It is described as having a cosmopolitan distribution, occurring in lakes and ponds of different trophic status and with different physical and chemical characteristics. Two molecular studies have been published with contrasting results regarding genetic diversity within this diatom. Based on allozyme electrophoresis of three polymorphic enzyme loci, Soudek & Robinson (1983) did not detect a single genetically different isolate within individual North American and European lakes, suggesting that *A. formosa* is a pure clonal species or at least genetically highly homogeneous. This is contradicted by the findings of De Bruin *et al.* (2004) who applied AFLP and RAPD methods and found each single isolate to be genetically unique within two *A. formosa* populations from the Dutch lakes Maarsseveen and Vinkeveen (De Bruin 2006). The latter suggests that sexual recombination (if only occasionally) in this species is likely to occur, despite the fact that sexual reproduction in *Asterionella* has never been observed. These contrasting results most likely reflect the resolution power of the different markers but may also reflect truly different levels of genetic variation in different lake populations. Nevertheless, both studies found considerable genetic differentiation among

populations of different lakes, but Soudek & Robinson (1983) did not observe a clear geographical pattern. Neighbouring lakes, even when connected by streams, were genetically not more similar to each other than distant lakes (Soudek & Robinson 1983). However, a drawback of this study was that isolates were collected in different years. Ideally, individuals sampled for estimation of population genetic structure should belong to the same generation or cohort, because allele frequencies vary not only over space, but also over time (Balloux & Lugon-Moulin 2002).

Concluding, the objective of this study was to investigate the genetic diversity and reproductive mode of the ubiquitous freshwater diatom *A. formosa* with high-resolution microsatellite and AFLP markers and to determine the spatial population structure during a single phytoplankton spring bloom.

Methods

Sampling design and isolation of clonal lineages

All isolates of *Asterionella formosa* were collected during the spring bloom of 2010 (January–March) with the exception of Lake Zwemlust (NL-Z) (February and April 2009) and Lake Greifen (CH-GS) from which we collected isolates in December 2009 and March 2010. In total, we sampled 11 lakes following a hierarchical spatial sampling design to assess genetic diversity and to reveal population structure on a (i) inter-regional scale (Swiss lakes vs. Dutch lakes), (ii) regional scale [seven lakes within Switzerland (CH) and four lakes within the Netherlands (NL)] and (iii) local scale (within lakes). In Switzerland, we also included lakes that were connected with each other, that is Lake Baldegg (CH-BS) which is connected to Lake Hallwil (CH-HW) by the river Aabach and the highly connected Upper and Lower Lake Zurich (CH-ZSO and CH-ZSU). Within Lake Lucerne (CH-VWS), we sampled seven different basins of this complex lake system. We divided the isolates of the seven different locations into three nominal populations (CH-VWSA, B and C), representing the most separated basins: Lake Alpnach (CH-VWSA); Horw Bay, Lake Küssnacht, Kreuztrichter and Weggis/Vitznau basin (CH-VWSB); and Gersau basin and Lake Uri (CH-VWSC) (Bührer & Ambühl 2001). Table 1 summarizes the sampled lakes and basins and their main characteristics. At each sampling location, a 10-m-depth integrated plankton sample was collected with a plankton net (30 μm mesh size). Within 24 h of collection, single colonies were isolated from a diluted plankton sample with an elongated glass pipette to obtain clonal cultures. Single colonies were transferred into individual wells of a 48-well plate containing 2 mL CHU 10 medium (Stein 1973) and maintained in a climate chamber at 14 °C on a 14:10-h light:dark cycle of 35 μmol photons/m^2/s for further growth. After 3 weeks of growth, each unialgal *A. formosa* isolate was transferred into Erlenmeyer flasks containing fresh CHU 10 medium to obtain sufficient material for DNA extraction.

DNA extraction

Algal material (1.5 mL) was collected during exponential growth in 2.5-mL Eppendorf tubes and was centrifuged to obtain an algal pellet. Supernatant was removed, and DNA was extracted following the HotSHOT DNA extraction protocol as described by Montero-Pau *et al.* (2008) using 20 μL of each buffer. As this was to our best knowledge the first time that the HotSHOT DNA extraction approach was used for diatoms and different DNA extraction methods may affect genotyping reproducibility, we performed a preliminary quality test. We compared the HotSHOT DNA extraction method and the Qiagen Plant Tissue Mini Kit with respect to the reproducibility of the genotype profile of four culture strains of *A. formosa* and obtained exactly the same genotyping profiles for all four strains with peaks of comparable quality.

Microsatellites development and genotyping

A microsatellite-enriched library was constructed by Ecogenics GmbH (Zurich, Switzerland) from size-selected genomic DNA ligated into SAULA/SAULB linker (Armour *et al.* 1994) and enriched by magnetic bead selection with biotin-labelled $(CT)_{13}$ and $(GT)_{13}$ oligonucleotide repeats (Gautschi *et al.* 2000a,b). Of 528 recombinant colonies screened, 213 gave a positive signal after hybridization. Plasmids from 155 positive clones were sequenced, and primers were designed for 16 microsatellite inserts. From these 16 microsatellites, nine polymorphic markers that could reliably be scored were selected. Details on the nine markers are summarized in Table 2. These nine markers were combined into two sets of multiplex (MP) PCRs (MP1: Ast01, Ast02 and Ast04; and MP2: Ast03, Ast05, Ast10, Ast13 and Ast14). Multiplex PCR amplification was performed in a 10 μL reaction volume containing 10 ng of DNA, 5 μL Multiplex PCR Master Mix (Qiagen), 0.3 μM of forward and reverse primers each and double-distilled, ultrapure water. Forward primers were fluorescently labelled with FAM, NED, PET or VIC fluorescent dyes (Applied Biosystems) (Table 2). The following amplification protocol was used on a Biometra thermal cycler (Whatman Biometra): 30 cycles of 30 s at 94 °C, 1 min 30 s at 60 °C (MP1) or 56 °C (MP2), respectively, 1 min at

Table 1 Summary characteristics of the sampled lakes and basins

Lake (abbreviation)	Country	LatitudeLongitude	Origin	Surface are (km^2)	Max depth (m)	Trophic status	Sampling date	Abundance *Asterionella formosa*
Lake Baldegg (BS)	Switzerland	47°11′47.69″N 8°15′45.17″E	Natural	5.2	66	Mesotrophic	15 March 2010	High
Lake Halwill (HW)	Switzerland	47°16′50.59″N 8°13′01.79″E	Natural	10.2	47	Mesotrophic	23 March 2010	Low
Lake Greifen (GS)	Switzerland	47°20′58″ 8°40′49″	Natural	8.45	32	Eutrophic	18 December 2009/24 March 2010	Moderate
Lake Rot (RS)	Switzerland	47°04′10.56″N 8°18′49.86″E	Natural	0.46	16	Meso-eutrophic	18 March 2010	High
Lower Lake Zurich (ZSU)	Switzerland	47°18′25.24″N 8°34′37.8″E	Natural	65	145	Mesotrophic	18 March 2010	High
Upper Lake Zurich (ZSO)	Switzerland	47°12′33.29″N 8°49′55.62″E	Natural	20	48	Oligo-mesotrophic	25 March 2010	Low
Lake Lucerne	Switzerland		Natural					
Lake Alpnach (VWSA)		47°07′47.70″N 8°19′25.04″E		4.76	35	Oligotrophic	24 February 2010	Low
Horw Bay, Lake Küssnacht, Kreuztrichter, Weggis/Vitznau basin (VWSB)		47°01′25.80″N 8°18′58.84″E		57	151	Oligotrophic	24 February 2010	Moderate
Lake Uri and Gersau basin (VWSC)		46°57′9.27″N 8°36′15.16″E		52	214	Oligotrophic	24 February 2010	Moderate
Fort Vechten (FV)	The Netherlands	52°05′81.31″N 5°16′42.48″E	Artificial 18th century	0.08	4	Eutrophic	2 March 2010	Unknown
Lake Maarsseveen (MSV)	The Netherlands	52°14′28.28″N 5°.08′57.11″E	Artificial 1965	0.7	32	Oligo-mesotrophic	26 January/25 March 2010	High
Lake Vinkeveen (VKV)	The Netherlands	52°23′52.12″N 4°96′15.98″E	Artificial 19th century	0.6	Shallow	Eutrophic	2 March 2010	Unknown
Lake Zwemlust (Z)	The Netherlands	52°19′30.40″N 5°00′73.13″E	Artificial 1921	0.015	2.5	Hypertrophic	25 February/21 April 2009	Unknown

Table 2 Summary of locus characteristics, PCR details and levels of variability of nine polymorphic microsatellite loci in the freshwater diatom *Asterionella formosa* from 14 locations in Switzerland and the Netherlands (n = 224)

	Locus	Dye	Primer sequence (5′→3′)	Repeat motif	T_a	Size	N_a
Multiplex 1	Ast01	NED	F CTACCGATAGCAGCCCAAG R ACGAATCAAGAAGCCGAGAC	$(AG)_{14}$	60	108–128	10
	Ast02	FAM	F CTGTCCTGCCTAACGGATG R GGAGCATGGTACACCCAAAG	$(AC)_{23}$	60	112–156	18
	Ast04	PET	F TGCATACAACTTGGCCCTTAC R TGTCCGAGTTGTTGTGTTCC	$(TC)_{12}$	60	85–101	13
	Ast08	VIC	F TTTCAATTTGACGTTTCCTCAC R AACGGAACAACAGCTTCTGG	$(AC)_{13}$	60	102–114	12
Multiplex 2	Ast03	FAM	F CCGTTACAACCCATGATACG R TCCCTTTTTGTTTGATTCGAC	$(GA)_{13}\ldots(GA)_6$	56	60–116	18
	Ast05	FAM	F CGGTCCAATGGTAAGACTCC R ATTGGAAAGCCGCAGTGTC	$(CA)_5(TA)_5\ldots(CA)_{17}$	56	196–248	13
	Ast10	NED	F TGATGGTCATCGAACTGATTG R CAGAGGCCAGTGGAGAAATG	$(GT)_7(GA)_{12}$	56	235–245	5
	Ast13	PET	F TTCTTGCGTGTCAAGAATGC R CAAATGGAATGGTTGGGTTC	$(TG)_{11}$	56	154–166	8
	Ast14	NED	F CTCGCTTGGGCATTCTGTAG R TGATGACGCTTGGTCTCAAC	$(TG)_{10}$	56	150–160	5

T_a, annealing temperature, N_a, number of alleles.

72 °C. Before the first cycle, a prolonged denaturation step (95 °C for 15 min) was included and the last cycle was followed by a 30-min final extension step at 72 °C. PCR products were 1:20 diluted with double-distilled, ultrapure water and, subsequently, 10 μL HiDi formamide and 0.25 μL size standard (GeneScan-500 LIZ; Applied Biosystems) were mixed with 1 μL of diluted PCR product. PCR products were separated on an ABI PRISM® 3130 XL Genetic Analyzer 16 Capillary system and analysed using GENEMAPPER version 3.7 (Applied Biosystems). Genotyping (i.e. PCR, electrophoresis and allele scoring) was repeated three times for 25% of the samples to assure genotype accuracy and reproducibility. The occurrence of null alleles and PCR amplification bias against large alleles was assessed using MICROCHECKER (Van Oosterhout *et al.* 2004).

In addition, we analysed a subset of isolates (91 of 224 isolates) with AFLP markers. AFLP analysis was performed by Keygene® (Wageningen, The Netherlands) using the same four *Eco*RI/*Mse*I AFLP primer combinations as reported in De Bruin *et al.* (2004): (i) Eco + GA and Mse + AT, (ii) Eco + GA and Mse + CC, (iii) Eco + GA and Mse + CG and (iv) Eco + GC and Mse + AC.

Population genetic structure

Cluster analysis based on microsatellite genotyping. Population structure was first assessed with a factorial correspondence analysis (FCA), performed on isolates with a unique multilocus genotype (MLG), as implemented in GENETIX version 4.05 (Belkhir *et al.* 1996–2004). To test the robustness of the clustering, pattern analyses were repeated after removing loci that showed evidence for null alleles and rare alleles. To further identify clusters of genetically related individuals and to obtain probabilities of assignment for individuals to a cluster (or clusters in case of admixture), two different approaches were used. First, a Bayesian clustering analysis was conducted with the software program STRUCTURE 2.3.3 (Pritchard *et al.* 2000). A minimum of 10 independent runs with unique MLGs were performed for each genetic cluster (K) ranging from 1 to 14 (reflecting the number of sample locations). The model assumed admixture, correlated allele frequencies and noninformative priors. A burn-in and run length of 100 000 iterations each was selected. The ΔK method (Evanno *et al.* 2005), implemented in the program STRUCTURE HARVESTER (Earl & vonHoldt 2012), was used to infer optimal number of clusters representing the data. STRUCTURE assumes Hardy–Weinberg equilibrium (HWE) and linkage equilibrium. As clonal organisms often fail to meet this assumption, the multivariate method discriminant analysis of principal components (DAPC), available in the ADEGENET package (Jombart 2008) for the R software (R Development Core Team), was used. This method does not rely on a particular population genetics model and is therefore free of assumptions on HWE or linkage disequilibrium (Jombart *et al.* 2010). The number of clusters was assessed using the function *find.clusters*, which runs successive K-means clustering with increasing numbers of clusters (K). Selection of the 'optimal number' of clusters was based on visual inspection of

the plots showing associated Bayesian information criterion (BIC) values for each number of clusters. In addition, the criterion 'goodfit' was used to infer the optimal number of clusters.

Pairwise F_{ST} values for each possible pairwise combination of populations were estimated based on 1000 permutations in ARLEQUIN (Excoffier *et al.* 2005). F_{ST} and related statistics have the drawback that they may underestimate population differentiation when polymorphism is high, which is usually the case for microsatellite markers (Heller & Siegismund 2009). Therefore, we also estimated Jost's *D*, an alternative measure of genetic differentiation which is not dependent on marker variability (Jost 2008). Locus-specific calculations of D_{EST} were performed using the software program SMOGD (Crawford 2010). The overall D_{EST} for each pairwise population comparison was calculated as the arithmetic mean across loci.

Cluster analysis based on AFLP genotyping. Amplified fragment length polymorphisms were analysed using Gaussian finite mixture models from the MCLUST package in R (Fraley 2002; Fraley *et al.* 2012). Different mixture probability distributions were built from the AFLP data. Each distribution represents a cluster consisting of data points themselves having certain probability belonging to the respective distribution. Model parameters were estimated by the expectation-maximization algorithm and tested for 3–35 clusters, and the different models were compared via BIC. The results were visualized using the dimension reduction method for model-based clustering (Scrucca 2010).

Genotypic richness and genetic diversity analysis

Individuals with identical microsatellite-based MLGs were detected with the program GENCLONE version 2.1. (Arnaud-Haond & Belkhir 2007). To assess whether identical MLGs could be considered as a part of the same asexual lineage, the probability that two individuals sharing the same MLG originate from different sexual reproductive events (P_{sex}) was estimated using MLGSIM2.0 (http://www.rug.nl/research/theoretical-biology/downloads), an updated version of MLGSIM (Stenberg *et al.* 2003). The program utilizes a simulation approach (1000 simulations) to assess statistical significance of P_{sex} values. Calculations were performed under the assumption of random mating (MODEL = HWE) as well as inbreeding (MODEL = FIS), and in each case, allele frequencies were estimated either from the entire sample (FREQUENCY = SAMPLE) or a subset of individuals with every MLG included only once (FREQUENCY = MLG), resulting in a total of four analyses per data set. Samples were grouped and analysed according to lake (or sub-basin) of origin, cluster assignment and lake by cluster assignment. Moreover, for each single population of these groups, genotypic richness (*R*) was estimated by ($R = G-1/N-1$), where *G* is the number of distinct MLGs and *N* is the number of sample units (Dorken & Eckert 2001). Hence, *R* ranges between 0 (monoclonal population) and 1 (each individual represents a genetically unique MLG). In order to directly compare lake populations with different sample size, we also performed a rarefaction analysis to calculate genotypic richness using the ARAREFACTWIN software (v. 1.3; S. Holland, Stratigraphy Lab, University of Georgia, Athens; www.uga.edu/~strata/software/). The smallest sample size (ZSO; $n = 10$) of the 13 lakes assayed was used to compare genotypic richness estimates between lakes. We further measured genetic diversity as allelic richness and expected heterozygosity as formulated by Nei (1987). We also estimated allelic richness with a rarefaction approach using the program HP-RARE (Kalinowski 2004) which fixes *n* as the smallest number of individuals typed for a locus in a sample. As two markers showed an elevated number of amplification failure within some populations, we performed the analysis with seven markers to avoid the fact that null amplification results in an underestimation of allelic richness. Weir & Cockerham's (1984) F_{IS} values and Nei's unbiased expected heterozygosity (H_e) and observed heterozygosity (H_o) were calculated with GENETIX version 4.05 (Belkhir *et al.* 1996–2004) for each locus. Hardy–Weinberg exact tests were performed using GENEPOP version 4.0 (Rousset 2008) assuming that the alternative hypothesis H_1 is a heterozygote deficiency. Average F_{IS} over all loci and linkage disequilibrium was calculated in ARLEQUIN (Excoffier *et al.* 2005). Linkage was estimated between all pairs of loci, with 10 000 dememorizations, 100 batches and 5000 iterations per batch. Critical significance levels were adjusted for multiple comparisons using the sequential Bonferroni correction (Rice 1989).

Results

Cryptic population structure

Factorial correspondence analysis performed on the clone-corrected data set resulted in the partitioning of individuals into three distinct genetic clusters (Fig. 1). The largest cluster (FCA-A) represented a mix of individuals from all sampled lakes and basins. The second largest cluster (FCA-B) contained only individuals from Swiss lakes except CH-RS and CH-VWSC, and the smallest cluster was only represented by CH-RS isolates (FCA-C). This pattern was robust towards removing loci that showed evidence for null alleles and rare alleles (data not shown).

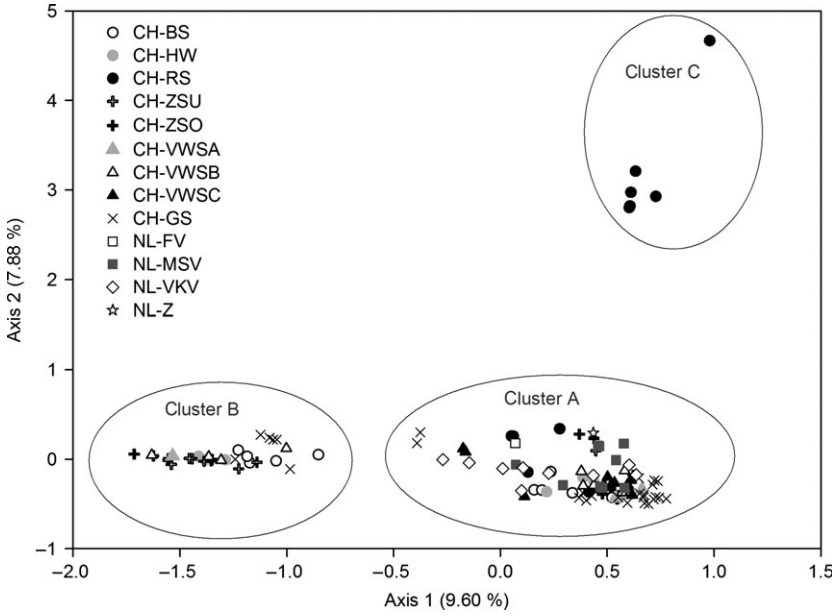

Fig. 1 Factorial correspondence analysis (FCA) including only unique multilocus genotypes (MLGs) of all *Asterionella formosa* isolates. Only the first two axes are shown.

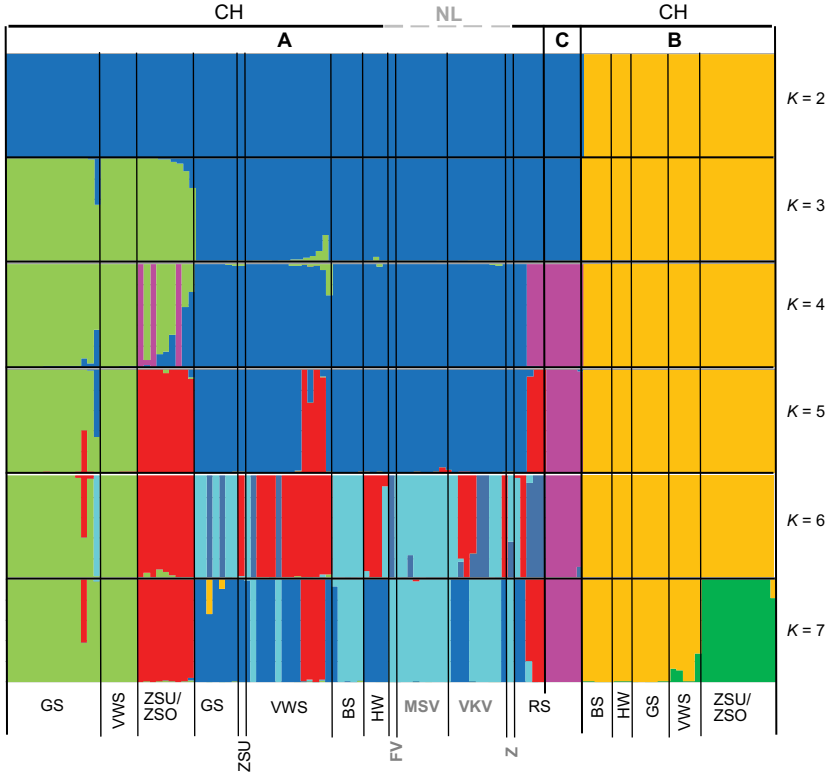

Fig. 2 Clustering of all unique multilocus genotypes (MLGs) of *Asterionella formosa* by discriminant analysis of principal components (DAPC), represented for different number of K ranging from $K = 2$ to $K = 7$. Each individual is depicted by a vertical bar that is partitioned into coloured sections, with the length of each section proportional to the estimated membership of the isolate to each cluster.

The DAPC analysis discriminated the same major clusters as found in the FCA (Fig. 2) but also revealed additional substructure. Overall, cluster assignment and sampling location were incongruent. As BIC values decreased continuously with increasing number of K, it was not straightforward to determine the optimal number of K (Fig. S1, Supporting information). Therefore, results are presented for different numbers of K, ranging from $K = 2$ to $K = 7$ (including $K = 6$ determined as optimal value for K by the 'goodfit' criterion). With $K = 2$, DAPC discriminated two clusters, the smaller cluster exactly corresponding to FCA-B, comprising MLGs from Swiss lakes only, and the larger cluster containing MLGs corresponding to the two remaining FCA

clusters, FCA-A plus FCA-C. This large cluster was again subdivided when allowing for 3 and 4 clusters. From $K = 5$ on also the cluster with a subset of CH-RS isolates equivalent to the small FCA-C cluster was consistently differentiated (Fig. 2). The fact that this cluster, being highly distinct in the FCA, does not appear in the DAPC before allowing for five clusters is due to the low sample size of FCA-C (for randomly subsampled data sets with reduced sample sizes for FCA-A and FCA-B, a cluster corresponding to FCA-C is consistently discriminated at $K = 3$). With increasing values for K, additional clusters representing groups of individuals originating from different lakes appeared. Altogether, only two clusters exclusively contained individuals from a single lake, the cluster with a subset of CH-RS isolates and another cluster with a subset of CH-ZSO + ZSU isolates. Similar results were found in the STRUCTURE analysis. An analysis of ΔK values indicated $K = 3$ as most likely number of clusters differentiating clusters FCA-A and FCA-B similar to DAPC. The CH-RS isolates of the FCA-C cluster also appeared at a higher number of K ($K = 6$) and the distinct CH-ZSO + ZSU cluster at $K = 9$ (Fig. S2, Supporting information). The same issue of large sample size variation between clusters, as seen with DAPC, is probably also affecting the clustering results of STRUCTURE (Kalinowski 2011).

For calculating estimators of differentiation, we grouped our samples according to sampling location and clusters FCA-A, FCA-B and FCA-C. F_{ST} and D_{EST} values for comparisons between the different clusters within a single lake were much higher than those for comparisons between lakes but within the same cluster (Tables 3 and 4 see boxed areas). We are aware that for most cases the sample size of the lake subpopulations is low and care must be taken with interpreting F_{ST} values and their significance. However, for CH-ZSU + ZSO sample sizes of both clusters (FCA-A and FCA-B) were still acceptable ($N > 10$). F_{ST} and D_{EST} values between clusters in this lake were 0.555 and 0.689, respectively, whereas values between CH-ZSU + ZSO and CH-GS of the same cluster were only 0.122 (but still significant) and 0.098, respectively. Despite the small sample sizes, the general pattern was consistent over all comparisons. The average F_{ST} and D_{ST} over all 'within lake/between clusters' comparisons was 0.482 (±0.05 SD) and 0.600 (±0.12 SD), respectively, and was much higher than the average F_{ST} and D_{ST} over all 'between lake/within clusters' comparisons, 0.200 (±0.09 SD) and 0.177 (±0.11 SD), respectively.

The cluster analysis based on AFLP data for a subset of isolates showed patterns mostly congruent with the microsatellite-based analyses (Fig. 3). A consistent pattern across both markers showed that isolates

Table 3 F_{ST} values between the sample locations and between subgroups within sample locations

		B					C	A									
	F_{ST}	CH-BS	CH-HW	CH-GS	CH-ZSU+ZSO	CH-VWS	CH-RS	CH-BS	CH-HW	CH-GS	CH-RS	CH-ZSU+ZSO	CH-VWS	NL-FV*	NL-MSV	NL-VKV	NL-Z*
B	CH-BS	0.000															
	CH-HW	−0.043	0.000														
	CH-GS	0.108	0.157	0.000													
	CH-ZSU+ZSO	**0.386**	**0.296**	**0.420**	0.000												
	CH-VWS	0.179	0.139	0.169	**0.233**	0.000											
C	CH-RS	**0.711**	0.815	**0.681**	**0.763**	**0.776**	0.000										
A	CH-BS	0.436	0.475	**0.451**	**0.542**	0.500	**0.562**	0.000									
	CH-HW	0.449	0.467	0.449	**0.570**	0.503	0.574	0.186	0.000								
	CH-GS	**0.518**	**0.534**	**0.515**	**0.589**	**0.555**	**0.534**	**0.273**	0.160	0.000							
	CH-RS	0.381	0.407	**0.389**	**0.493**	0.406	**0.485**	0.203	0.168	**0.343**	0.000						
	CH-ZSU+ZSO	**0.443**	**0.482**	**0.471**	**0.555**	**0.508**	**0.485**	**0.234**	0.147	**0.122**	**0.216**	0.000					
	CH-VWS	**0.397**	**0.424**	**0.384**	**0.483**	**0.433**	**0.421**	**0.187**	**0.151**	**0.116**	**0.226**	**0.093**	0.000				
	NL-FV*	0.506	0.638	0.458	0.644	0.599	0.742	0.282	0.179	0.357	0.215	0.310	0.214	0.000			
	NL-MSV	**0.402**	**0.413**	**0.406**	**0.497**	**0.428**	**0.425**	0.099	**0.147**	**0.210**	**0.186**	**0.176**	**0.170**	0.182	0.000		
	NL-VKV	**0.302**	0.315	**0.312**	**0.422**	**0.358**	**0.392**	0.112	0.105	**0.211**	0.124	**0.141**	**0.110**	0.138	0.072	0.000	
	NL-Z*	0.595	0.716	0.580	0.715	0.704	0.763	0.320	0.227	0.400	0.287	0.355	0.296	0.120	0.195	0.218	0.000

Significant F_{ST} values in bold ($P < 0.0005$ after sequential Bonferroni correction).
*Only one multilocus genotype present; areas boxed with broken lines designate within and among comparisons of subgroups A and B; boxed values represent examples with sample sizes ≥9 (see Table 6 for sample sizes), and those highlighted in grey shows an example of high F_{ST} values between subgroups A and B within lake CH-ZSU + ZSO compared to much lower F_{ST} values between samples of different lakes belonging to the same subgroup A; prefix CH (lakes located in Switzerland); prefix NL (lakes located in the Netherlands); Lake Baldegg (BS), Lake Halwill (HW), Lake Greifen (GS), Lake Rot (RS), Lower and Upper Lake Zurich (ZSU + ZSO), Lake Lucerne basin A (VWSA), Lake Lucerne basin B (VWSB), Lake Lucerne basin C (VWS C), Fort Vechten (FV), Lake Maarsseveen (MSV), Lake Vinkeveen (VKV), Lake Zwemlust (Z).

Table 4 D_{ST} values between the sample locations and between subgroups within sample locations

		B					C	A									
	D_{EST}	CH-BS	CH-HW	CH-GS	CH-ZSU+ZSO	CH-VWS	CH-RS	CH-BS	CH-HW	CH-GS	CH-RS	CH-ZSU+ZSO	CH-VWS	NL-FV*	NL-MSV	NL-VKV*	NL-Z
B	CH-BS	0.000															
	CH-HW	0.003	0.000														
	CH-GS	0.038	0.125	0.000													
	CH-ZSU+ZSO	0.102	0.076	0.108	0.000												
	CH-VWS	0.074	0.026	0.076	0.038	0.000											
C	CH-RS	0.746	0.779	0.621	0.774	0.761	0.000										
A	CH-BS	0.450	0.531	0.571	0.483	0.616	0.720	0.000									
	CH-HW	0.446	0.486	0.547	0.559	0.601	0.786	0.200	0.000								
	CH-GS	0.545	0.647	0.724	**0.766**	0.848	0.709	0.252	0.069	0.000							
	CH-RS	0.358	0.498	0.409	0.417	0.353	0.537	0.306	0.160	0.387	0.000						
	CH-ZSU+ZSO	0.446	0.656	0.615	**0.688**	0.675	0.585	0.292	0.146	**0.098**	0.161	0.000					
	CH-VWS	0.514	0.690	0.464	**0.735**	0.718	0.669	0.284	0.196	**0.073**	0.245	**0.088**	0.000				
	NL-FV	0.294	0.448	0.303	0.435	0.375	0.707	0.088	0.047	0.262	0.021	0.343	0.045	0.000			
	NL-MSV	0.570	0.715	0.622	**0.632**	0.579	0.677	0.177	0.307	**0.253**	0.305	**0.271**	**0.338**	0.122	0.000		
	NL-VKV*	0.376	0.443	0.412	**0.488**	0.587	0.643	0.212	0.118	**0.208**	0.217	**0.199**	**0.184**	0.023	**0.154**	0.000	
	NL-Z*	0.647	0.706	0.934	0.805	0.958	0.779	0.328	0.133	0.411	0.260	0.350	0.129	N.A.	0.379	0.039	0.000

D_{EST} for each pairwise comparison was calculated as the arithmetic mean across loci.
*Only one multilocus genotype present; areas boxed with broken lines designate within and among comparisons of subgroups A and B; boxed values represent examples with sample sizes ≥9 (see Table 6 for sample sizes), and those highlighted in grey shows an example of high D_{EST} values between subgroups A and B within lake CH-ZSU + ZSO compared to much lower D_{EST} values between groups of different lakes belonging to the same subgroup A; prefix CH (lakes located in Switzerland); prefix NL (lakes located in the Netherlands); Lake Baldegg (BS), Lake Halwill (HW), Lake Greifen (GS), Lake Rot (RS), Lower and Upper Lake Zurich (ZSU + ZSO), Lake Lucerne basin A (VWSA), Lake Lucerne basin B (VWSB), Lake Lucerne basin C (VWS C), Fort Vechten (FV), Lake Maarsseveen (MSV), Lake Vinkeveen (VKV), Lake Zwemlust (Z).

originating from the same lake grouped in different clusters. This resulted in two very heterogeneous clusters (clusters depicted in blue and red colours in Fig. 3) that contained a mixture of isolates originating from several lakes spanning the whole geographical sampling range. Besides these heterogeneous clusters, also lake-specific clusters were identified. The first axis separated the isolates from NL-Z. The second axis separated five other lake-specific clusters. The CH-RS cluster corresponded to the microsatellite-based FCA-C cluster. The differentiation between the major clusters A and B was not as clear as with microsatellite markers. However, AFLP markers also differentiated a distinct CH-ZSU cluster that contained only isolates present in the FCA-B cluster (with the exception of a single isolate; ZSU23) and which also formed a separate group in DAPC and STRUCTURE. The rest of the isolates that were designated to cluster B in the microsatellite analysis formed a single cluster but included also some isolates from cluster A. The third axis positioned this mixed cluster between the CH-ZSU isolates of cluster B and the other groups of isolates which all belonged to cluster A.

Linkage disequilibrium and heterozygote deficiency

Significant linkage disequilibrium was found in five lake samples with highest proportions observed in CH-GS (67%), CH-RS (25%) and CH-ZSO (28%) (Table 5). All samples showed positive F_{IS} values across all loci and thus a significant heterozygote deficiency, except for CH-VWSC and NL-MSV from which the genotypic distributions did not deviate from HWE expectations (Tables 5 and S1, Supporting information). For all samples, F_{IS} values showed large variation across loci (Table S1, Supporting information). When individuals of each lake were separated into subgroups according to their cluster assignment (i.e. A, B and C), F_{IS} values decreased and became nonsignificant (Table 6). Linkage disequilibrium also generally decreased, except for the pooled CH-VWS basins (Table 6).

Genotypic and genetic diversity (lake and within-lake subpopulation level)

We detected 122 distinct MLGs among the entire 224 individuals. A total of 132 individuals were represented by 30 recurring MLGs comprising 2–20 individuals each. Considering the 182 individuals only that could be genotyped for all nine markers, we found 94 distinct MLGs. When excluding two markers (Ast 04 and Ast 03) with an increased number of amplification failure and all individuals with amplification failure in one of the other markers, we identified 96 distinct MLGs among 214 individuals. Most of the MLGs were specific

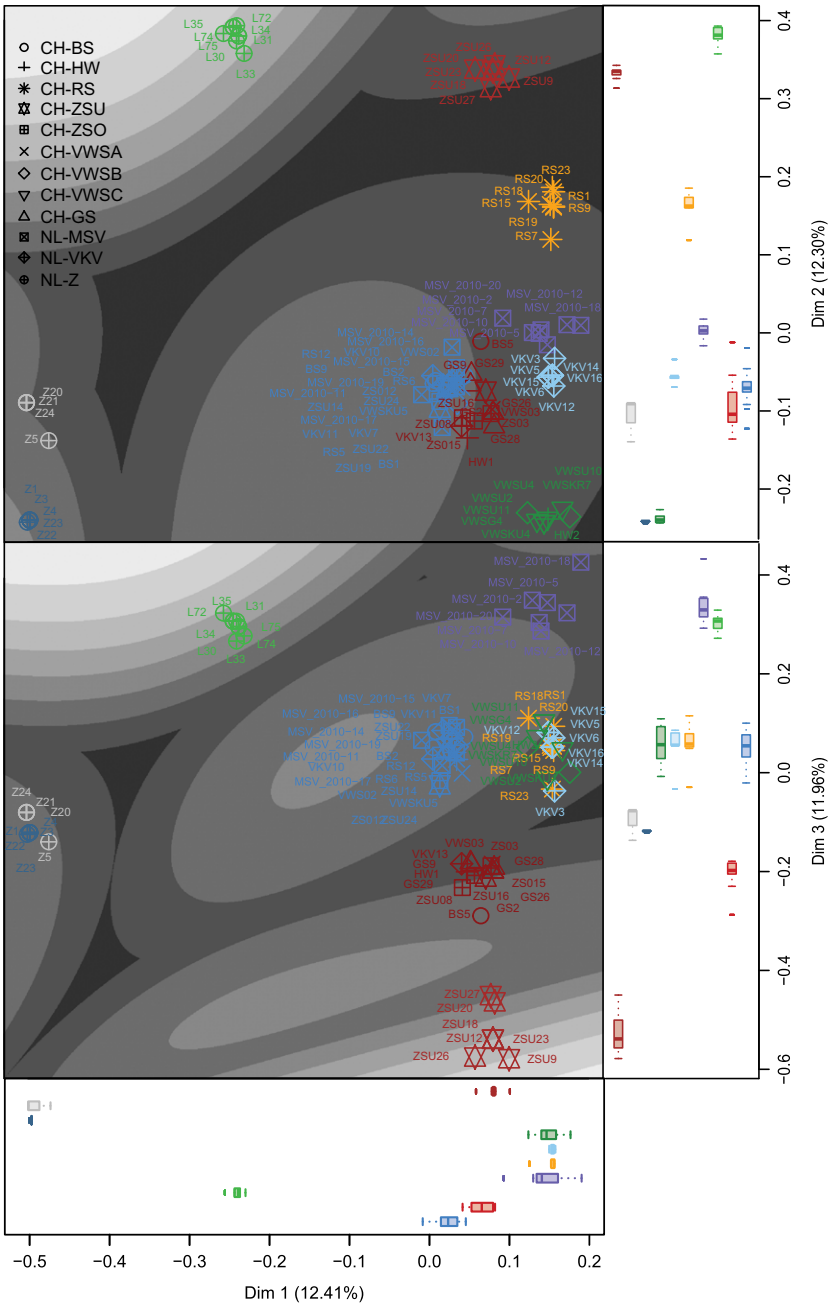

Fig. 3 Dimension reduction for the model-based clustering of the amplified fragment length polymorphism (AFLP) data. Shown are the first three axes (linear combinations of the original features) associated with the highest eigenvalues. They account for 36.7% of the clustering structure. Lake and cluster affiliations are depicted by different characters and colours, respectively.

for their lake of origin. One MLG was shared by the two unconnected but nearby lakes CH-ZSO and CH-HW. CH-ZSO and CH-ZSU also shared one MLG and two MLGs each were shared by the basins CH-VWSA and B and basins B and C, respectively. One MLG was found in all three basins of Lake Lucerne. P_{sex} values were generally very low and indicated clonal identity in most cases. The vast majority of P_{sex} values were significant for all four parameter combinations in the MLGSIM2.0 simulations: a data set comprising all MLGs 100%, data sets discriminating the three main clusters 89.65%, data sets partitioning lakes by clusters 81.25% and data sets based on lakes 96.67% of P_{sex} values were statistically significant ($P < 0.05$) (see Table S2, Supporting information). Genotypic richness was variable among lakes and ranged from 0 to 0.79 (Table 7). Based on rarefaction curves (Fig. 4), the CH lakes ZSO, ZSU, GS, HW, VWSB and VWSC had a similar degree of genotypic richness (lying within the 95% confidence intervals for the lake with the smallest sample size). The CH lakes RS, BS and VWSA and NL lakes MSV and VKV had lower genotypic richness. Extremely low genotypic richness was found in the two Dutch lakes NL-FV and NL-Z, for which microsatellite analysis

Table 5 Wright's inbreeding coefficient (F_{IS}), Hardy–Weinberg test statistics (HWE *P*-val) and linkage disequilibrium (LD)

Region	Sample	*N*	F_{IS}	HWE *P*-val	LD
CH	BS	10	0.27	***	0 (36)
CH	HW	7	0.32	***	0 (36)
CH	GS	28	0.34	***	24 (36)
CH	RS	9	0.39	***	7 (28)
CH	ZSU	12	0.34	***	10 (36)
CH	ZSO	8	0.29	***	0 (36)
CH	VWSA	5	0.18	*	0 (28)
CH	VWSB	12	0.23	**	3 (36)
CH	VWSC	9	−0.01	0.737	0 (36)
NL	FV	1	NA	NA	NA
NL	MSV	9	−0.11	0.823	0 (28)
NL	VKV	8	0.23	**	1 (36)
NL	Z	1	NA	NA	NA

HWE, Hardy–Weinberg equilibrium.
Linkage disequilibrium (LD): number of pairwise combinations of loci that were in significant genotypic disequilibrium after sequential Bonferroni correction (the number of tested combinations is in parentheses); for VWSA locus, Ast04 was removed due to missing data; monomorphic loci occurring in some samples were also excluded.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Table 6 Wright's inbreeding coefficient (F_{IS}), Hardy–Weinberg test statistics (HWE *P*-val) and linkage disequilibrium (LD) for each group separated according to the major clusters A, B and C

Region	Sample	*N*	F_{IS}	HWE *P*-val	LD
	Cluster A				
CH	BS	5	−0.19	0.88	0 (36)
CH	HW	4	0.04	0.44	0 (36)
CH	GS	22	0.003	0.25	8 (28)
CH	RS	5	0.10	0.16	0 (36)
CH	ZSU + ZSO	10	−0.07	0.86	4 (28)
CH	VWS	18	0.05	*	21 (36)
NL	MSV	9	−0.11	0.83	0 (36)
NL	VKV	9	0.23	**	1 (36)
	Cluster B				
CH	BS	5	0.26	*	0 (21)
CH	HW	3	−0.22	0.87	NA
CH	GS	6	0.05	0.11	0 (13)
CH	ZSU + ZSO	12	0.03	0.33	0 (15)
CH	VWS	5	0.11	0.13	NA
	Cluster C				
CH	RS	6	−0.41	1.00	NA

HWE, Hardy–Weinberg equilibrium; NA, not applicable.
Linkage disequilibrium (LD): number of pairwise combinations of loci that were in significant genotypic disequilibrium after sequential Bonferroni correction (the number of tested combinations is in parentheses); for VWSA locus, Ast04 was removed due to missing data; monomorphic loci occurring in some samples were also excluded.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Table 7 Genotypic and genetic diversity of *Asterionella formosa* within lakes

Sample	#*M*	*N*	*G*	*R*	AR*	SE_{AR}	H_o	Div
BS	9	22	10	0.43	3.64	0.45	0.49	0.67
HW	8	10	7	0.67	3.70	0.18	0.48	0.70
GS	8	35	28	0.76	4.19	0.52	0.45	0.68
RS	9	20	9	0.42	2.97	0.46	0.36	0.59
ZSU	9	22	12	0.52	3.15	0.29	0.38	0.58
ZSO	9	10	8	0.78	3.84	0.59	0.47	0.66
VWSA	8	13	5	0.33	3.28	0.39	0.55	0.67
VWSB	8	15	12	0.79	4.46	0.58	0.54	0.71
VWSC	8	13	9	0.67	4.41	0.46	0.69	0.68
FV	9	9	1	0.00	1.71	0.18	NA	NA
MSV	8	14	9	0.54	4.14	0.62	0.76	0.69
VKV	8	13	8	0.58	4.23	0.53	0.58	0.76
Z	9	20	1	0.00	1.71	0.18	NA	NA
Σ pop	7	214	96	0.45	10.12	1.52		
Σ pop	9	182	94	0.51	11.22	1.60		

NA, not applicable.
Number of markers used to calculate genotypic richness (#*M*), sample size (*N*), number of genotypes (*G*), genotypic richness (*R*), mean allelic richness based on seven markers (AR*), standard error of AR (SE_{AR}), observed heterozygosity (H_o), Nei's gene diversity 1987 (Div).

detected only a single genotype (Table 7, Fig. 4) (note that AFLP markers discriminated more genotypes but dissimilarity coefficients were very low). Gene diversity (Nei's unbiased expected heterozygosity) among lakes ranged from 0.58 to 0.76 (Table 7). Genotypic richness between subgroups (based on FCA and DAPC analysis) differed for some lakes (Table 8) but without showing a consistent pattern. Subgroups of clusters B and C, however, had systematically lower observed heterozygosity and gene diversity compared to the subgroups of cluster A (Table 8).

Discussion

Cryptic structure and hidden diversity

With this study, we aimed at investigating whether the cosmopolitan distribution of the freshwater pelagic diatom *Asterionella formosa* also translates into a homogeneously distributed gene pool or whether there is evidence for spatial genetic population structure. The former would be indicative of high gene flow (i.e. unlimited dispersal rates) and would thus support the 'everything is everywhere' hypothesis. In the latter case, spatially fragmented gene pools would be indicative of restricted gene flow, a necessary condition for allopatric speciation and biogeographical patterns to arise.

Our results revealed a complex and unexpected population structure in *A. formosa* with evidence for both

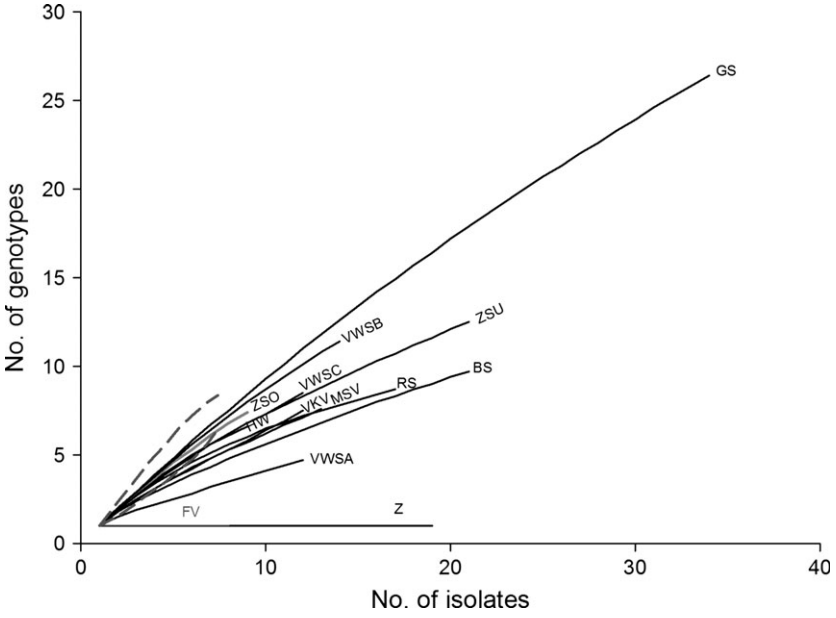

Fig. 4 Rarefaction curves (solid lines) show the genotypic richness of *Asterionella formosa* in each lake and lake basin. Dotted grey lines are the 95% confidence intervals for the lake with the smallest sample size (ZSO; $n = 10$), which was used to compare genotypic richness estimates between lakes.

Table 8 Genotypic and genetic diversity of *Asterionella formosa* within lakes and subgroups

Sample	#*M*	*N*	*G*	*R*	H_o	Div
Cluster A						
BS	9	16	5	0.26	0.69	0.58
HW	8	6	4	0.60	0.61	0.63
GS	8	26	23	0.88	0.49	0.49
RS	9	9	5	0.50	0.49	0.55
ZSU + ZSO	7	16	8	0.47	0.56	0.52
VWS	8	28	17	0.59	0.65	0.68
FV	9	9	1	0.00	NA	NA
MSV	8	14	9	0.54	0.76	0.69
VKV	8	13	8	0.58	0.58	0.76
Z	9	20	1	0.00	NA	NA
Cluster B						
BS	9	6	5	0.80	0.29	0.39
HW	8	4	3	0.67	0.29	0.24
GS	8	8	5	0.57	0.37	0.39
ZSU + ZSO	7	18	9	0.47	0.28	0.29
VWS	8	13	3	0.17	0.28	0.31
Cluster C						
RS	7	11	4	0.30	0.19	0.13

NA, not applicable.
Number of markers used to calculate genotypic richness (#*M*), sample size (*N*), number of genotypes (*G*), genotypic richness (*R*), observed heterozygosity (H_o), Nei's gene diversity 1987 (Div).

restricted and moderate to high gene flow at the same time. Different genetic markers (microsatellites and AFLPs) analysed with a variety of multivariate methods (FCA, DAPC, STRUCTURE, Gaussian finite mixture models) consistently revealed that genetic differentiation within lakes was much stronger than among lakes, even among geographically distant lakes. The presence of strong within-lake substructure was also confirmed by population genetic analysis. High heterozygote deficiency and high proportions of loci in significant linkage disequilibrium were observed when isolates were pooled together according to their lake of origin. Linkage across loci could be a sign for a dominant clonal reproductive mode. However, under such conditions one would also expect negative F_{IS} values, indicating an excess of heterozygotes due to the 'Meselson effect' (Halkett *et al.* 2005). Heterozygote deficiency could be the result of null alleles or inbreeding. Despite exclusion of loci showing evidence for null alleles from the analysis, F_{IS} values remained positive. High levels of inbreeding/selfing seem to be also unlikely because that should affect F_{IS} values across all loci in an equal manner (Conner & Hartl 2004), which was not the case in our data. Separating lake samples according to their assignment to the three main FCA clusters resulted in a considerable decrease of F_{IS} values and LD and deviations from HWE became insignificant in six of nine cases. Our results thus indicate that the main cause for deviations from Hardy–Weinberg was substructure (Wahlund effect) within lakes. Moreover, F_{ST} and D_{EST} values confirmed the strong within-lake divergence; that is, values were much higher between samples of the same lake than between samples of different lakes. Such a high divergence can only be achieved when gene flow between lakes is much higher than within lakes. The observed pattern suggests reproductive isolation between coexisting subgroups and hints towards the presence of cryptic species in *A. formosa*. Pappas & Stoermer (2001) already suggested that the genus

Asterionella necessitates taxonomic enquiry and they provide evidence for at least seven distinct valve shape variants of *A. formosa* from the Great Lakes based on morphometric data. The present study found F_{ST} values between sympatric subpopulations that are above or in a range (0.433–0.555) used to designate cryptic species within other groups, such as ascidians (F_{ST} values of 0.276) (Caputi *et al.* 2007), but also the marine diatoms *P. pungens* (F_{ST} values of 0.481) (Adams *et al.* 2009) and *Ditylum brightwelli* (F_{ST} values of 0.286) (Koester *et al.* 2010). In particular, the results of Adams *et al.* (2009) are very similar to our findings, that is two distinct groups of *P. pungens* from a single location in the Pacific Northwest that were more divergent from each other (F_{ST} values of 0.481) than either was from a sample from the North Sea (F_{ST} values of 0.274 and 0.352). They also interpreted their finding as a result of the presence of cryptic species. Unfortunately, we do not have sequences of species-level phylogenetic markers like the internal transcribed spacer region or the mitochondrial cytochrome *c* oxidase subunit 1 (cox1) available. However, a follow-up study should compare sequence divergence of such species-level markers with the observed subpopulations. This will allow us to place the observed differentiation in an evolutionary context, that is how far divergence has proceeded.

Sympatric reproductive isolation and cryptic species have recently been reported for other diatoms (e.g. Mann 1999; Amato *et al.* 2007; Poulickova *et al.* 2010). One characteristic for diatom life history is the progressive cell size reduction during asexual growth with a reconstitution of cell size through sexual reproduction when a certain minimum cell size is attained (Lewis 1984). The simultaneous occurrence of different size classes within a lake has been observed in many freshwater diatoms, including *A. formosa* (Mann 1988), although in the case of *Asterionella* this could not be linked to sexual reproduction and auxospore formation. Genotypes with different cell sizes may attain their sexual phase at different time points, thereby exchanging genes only with a certain subset of genotypes. Such asynchronous reproduction periods could generate coexisting populations with strong genetic differentiation (Santos *et al.* 2013). Another mechanism that may produce reproductive isolation and cryptic speciation is genome duplication (Koester *et al.* 2010). This was found to have occurred in two genetically different sympatric populations of the marine diatom *D. brightwellii* where the population with a larger mean cell size was found to have double the amount of DNA content, indicative for genome duplication (Rynearson *et al.* 2006; Koester *et al.* 2010). Interestingly, F_{ST} values between those populations were 0.286 and thus much lower than F_{ST} values found between sympatric populations of different clusters in the present study. We measured frustule length of a subset of isolates to check whether we could also detect any morphological differences. For CH-ZSU and ZSO, we found two different size classes that corresponded to the two genetically distinct clusters (Fig. S3, Supporting information). This suggests that asynchronous reproduction periods or genome duplication could be a potential explanation for the observed subdivision in CH-ZSU + ZSO. However, the lack of such a clear pattern in the rest of the lakes suggests that other or additional mechanisms may also be involved. Interestingly, an early study by Ruttner (1937) described the presence of two 'ecotypes' of *A. formosa* in alpine lakes on the basis of their frustules length and vertical distribution. Moreover, limnological studies on Lake Zurich report the occurrence of *Asterionella gracillima* until 1940, whereas later the species was renamed to *A. formosa*. Morphologically, the two species differ in their valve shape: isopolar (*A. gracillima*) vs. heteropolar (*A. formosa*). We did not observe this morphological difference in our study. However, it is possible that this morphological trait is variable and overlapping which could have induced the confusion about the species nomination in Lake Zurich.

It is intriguing that the subgroup division into clusters A and B of the Lake Zurich strains was also depicted by AFLP markers, whereas this division was less clear for the other lakes in the AFLP analysis. It is possible that within the different lakes the two lineages are in different stages of divergence with Lake Zurich subgroups showing strongest divergence, as supported by the highest intralake F_{ST} values. An explanation for this observation may be sympatric speciation or secondary contact with gene flow of varying degree due to differences in the timing of secondary contact or strength of ecological selection and reproductive isolation between lakes. Such scenarios are expected to create highly heterogeneous patterns of divergence across the genome (Noor & Bennett 2009; Nosil *et al.* 2009; Feder *et al.* 2012; Seehausen *et al.* 2014). AFLPs cover a broader range of the genome, whereas microsatellite loci are restricted to a few loci, and therefore, differences between these markers are to be expected in the case of heterogeneous genomewide divergence.

Co-occurrence of several closely related and morphologically similar lineages is paradoxical given conventional niche theory (Hutchinson 1961). Subgroups were not associated with trophic state because subgroups of both cluster A and cluster B were present in oligotrophic as well as eutrophic lakes. However, other environmental factors and traits which we did not measure could be involved. Parasites, for example, have been shown to drive genetic divergence in their host (Karvonen & Seehausen 2012). Sonstebo & Rohrlack (2011) found

indications for chytrid parasitism being the driving force for population subdivision and coexistence of different *Planktothrix* chemotypes. In this context, it is interesting to note that *Asterionella* is also a well-known host for chytrid parasites (Van den Wyngaert *et al.* 2014) and chytrid infections were often observed in our samples (S. Van den Wyngaert, personal observation, not quantified). This is an intriguing hypothesis that deserves further investigation. Alternatively, it has been shown that for protist communities Hubbell's neutral theory, based on stochastic demographic and dispersal events from a large species pool, can also explain species coexistence (Dolan *et al.* 2007). Moreover, it has been observed that cryptic species, even with specialized ecophysiological characteristics, may have widely overlapping spatial distributions (Vanelslander *et al.* 2009).

Clonal and genetic diversity

The fact that we frequently found identical MLGs and statistically significant P_{sex} values (Table S2, Supporting information) is in agreement with the known predominantly asexual mode of reproduction in the freshwater diatom *A. formosa*. In fact, sexual reproduction has so far never been observed in this diatom species. This may explain the rather low clonal diversity of *A. formosa* ($R = 0.51$ averaged over all samples) observed in this study compared to the majority of microsatellite-based studies on marine protists (Rynearson & Armbrust 2000; Evans *et al.* 2004; Iglesias-Rodriguez *et al.* 2006; Masseret *et al.* 2009; Lowe *et al.* 2010). It seems unlikely that the nature of the microsatellite markers used is causal for this difference as all loci showed a moderate (five alleles) to high (18 alleles) degree of polymorphism. The low average R value partly results from the large variation of clonal diversity among lakes. Two Dutch *A. formosa* populations showed extremely low diversity; that is, they were represented by a single clone. These populations reflect the strong clonality of this species as observed by Soudek & Robinson (1983). The reasons for this extreme low genotypic richness could further be due to both historical and morphological characteristics of these lakes. Both are artificial, isolated, small and shallow lakes. In the parthenogenetic water flea *Daphnia*, it has been shown that populations inhabiting smaller water bodies have lower clonal diversity, mainly due to smaller population sizes causing increased genetic drift (Vanoverbeke *et al.* 2007). Moreover, NL-Z was subjected to a recent biomanipulation event in 1989 during which the whole lake was drained and allowed to refill with groundwater (Gulati & van Donk 2002). Founder effects could play an additional role for the very low genetic diversity within this lake. However, even after exclusion of these two lakes with extreme low diversity, the average clonal diversity remained much lower than in other microsatellite-based studies.

To our best knowledge, only one other microsatellite study exists on genotypic diversity of a freshwater protist (benthic diatom *Sellaphora capitata*) and this study also reported high clonal diversity with 89% of the isolates being unique MLGs (Evans *et al.* 2009). It is difficult to estimate whether the lower clonal diversity observed in *A. formosa* is species specific or results from inherent differences between freshwater lake and marine habitats. The generally higher values for clonal diversity found by many studies might also be the result of different sampling procedures. In most studies, isolates from different time points (sometimes different years) are pooled together, whereas in the present work, isolates originate from a single sample, increasing the chance of isolating identical clones that are the result of asexual bloom development. In the marine diatom *D. brightwellii*, however, clonal diversity remained very high even during a spring bloom event (Rynearson & Armbrust 2005).

When comparing clonal diversity values for *A. formosa* from studies that used different markers, we can conclude that our results fall in between the two extreme findings of the allozyme study of Soudek & Robinson (1983) (no diversity within lakes) and the AFLP study of De Bruin *et al.* (2004) (each isolate genetically unique). The discrepancy between the levels of clonal diversity between studies is most probably due to the different markers used (Nybom 2004). Using microsatellite markers, we detected identical MLGs that belonged to the same clonal lineage in all lakes (including NL-MSV studied by De Bruin *et al.* 2004) with high probability. However, our AFLP analysis on a subset of the samples also differentiated much more unique genotypes than microsatellite markers (but also here clones occurred) and was as such in agreement with the results of De Bruin *et al.* (2004) and Gsell *et al.* (2013). An explanation for the discrepancy between both markers may be again the fact that AFLPs cover a broader part of the genome; that is, many loci are analysed compared to a few microsatellite loci.

When substructure was not considered, the average gene diversity (expected heterozygosity) was 0.67 and comparable with other diatoms. However, large variation in gene diversity was found between the subgroups with lake subgroups from the B and C clusters showing substantially lower gene diversity which may indicate bottlenecks or higher population turnover (e.g. Walser & Haag 2012).

Conclusion

Our results on the genetic diversity and population structure patterns in the common freshwater diatom

A. formosa showed surprising and interesting patterns. The strong genetic divergence within lakes raises the question of cryptic speciation in *A. formosa* and asks for more thorough investigations of species limits within this genus. This will require a combination of several tools. (i) Molecular analysis with mitochondrial markers such as *cox1* and *cob* genes has been successfully discriminated cryptic species in other protist studies (Alverson 2008). (ii) Fine-grained morphological and nuclear cytological investigations may reveal subtle differences between clusters (Poulickova *et al.* 2010). (iii) Physiological studies may help to clarify whether the observed neutral genetic divergence also reflects adaptive divergence as ecological differences may affect gene flow and therefore lead to genetic divergence (Gsell *et al.* 2012). (iv) Ideally, mating experiments should be performed to detect reproductive isolation between members of the different genetic groups (Mann 1999).

Over the past 20 years, an increasing number of cryptic and/or pseudo-cryptic microalgae species have been discovered within traditionally described cosmopolitan morphospecies (Beszteri *et al.* 2007; Trobajo *et al.* 2009; Poulickova *et al.* 2010; Smayda 2011). Moreover, the question whether dealing with large intraspecific genetic differentiation or the presence of cryptic species is becoming a recurrent issue in population genetic studies of diatoms and protists in general, which poses a challenge for the 'everything is everywhere' debate (Boenigk *et al.* 2012). *Asterionella formosa* is ubiquitous and one of the dominant spring bloom phytoplankton species in many temperate lakes. The hidden diversity found in this study is expected to have implications for the further use of *A. formosa* in biogeographical studies, ecological monitoring programmes and conservation and ecological studies in general (Mann 2010).

Acknowledgements

We thank Riet Schocher and Jean-Claude Bernegger from the Environment and Energy Competence Centre (uwe), Canton of Lucerne (Switzerland), for offering the facilities to sample Lake Lucerne, Lake Rot and Lake Baldegg. We also thank Alena Gsell, Jukka Jokela, Kirstin Kopp and Luc De Meester for helpful input and discussions. This research was funded by the ETH Board (CCES7 GEDIHAP). Additional support to Markus Möst was provided by a SNF Early PostDoc. Mobility Fellowship (P2EZP3_148773).

References

Adams NG, Trainer VL, Rocap G, Herwig RP, Hauser L (2009) Genetic population structure of *Pseudo-nitzschia pungens* (Bacillariophyceae) from the Pacific Northwest and North Sea. *Journal of Phycology*, **45**, 1037–1045.

Alpermann TJ, Beszteri B, John U, Tillmann U, Cembella AD (2009) Implications of life-history transitions on the population genetic structure of the toxigenic marine dinoflagellate *Alexandrium tamarense*. *Molecular Ecology*, **18**, 2122–2133.

Alverson AJ (2008) Molecular systematics and the diatom species. *Protist*, **159**, 339–353.

Amato A, Kooistra WHCF, Levialdi Ghiron JH *et al.* (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*, **158**, 193–207.

Armour JAL, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics*, **3**, 599–605.

Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, **7**, 15–17.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

Beijerinck MW (1919) *De infusies en de ontdekking der backteriën*. Jaarboek van de Koninklijke Akademie v. Wetenschappen Müller, Amsterdam.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Beszteri B, John U, Medlin LK (2007) An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. *European Journal of Phycology*, **42**, 47–60.

Boenigk J, Ereshefsky M, Hoef-Emden K, Mallet J, Bass D (2012) Concepts in protistology: species definitions and boundaries. *European Journal of Protistology*, **48**, 96–102.

Bührer H, Ambühl H (2001) Lake Lucerne, Switzerland, a long term study of 1961–1992. *Aquatic Sciences*, **63**, 432–456.

Caputi L, Andreakis N, Mastrototaro F, Cirino P, Vassillo SP, Sordino P (2007) Cryptic speciation in a model invertebrate chordate. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 9364–9369.

Casteleyn G, Chepurnov VA, Leliaert F *et al.* (2008) *Pseudo-nitzschia pungens* (Bacillariophyceae): a cosmopolitan diatom species? *Harmful Algae*, **7**, 241–257.

Casteleyn G, Evans KM, Backeljau T *et al.* (2009) Lack of population genetic structuring in the marine planktonic diatom *Pseudo-nitzschia pungens* (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. *Marine Biology*, **156**, 1149–1158.

Casteleyn G, Leliaert F, Backeljau T *et al.* (2010) Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 12952–12957.

Conner JK, Hartl DL (2004) *A Primer of Ecological Genetics*. Sinauer Associates, Sunderland, Massachusetts.

Crawford NG (2010) smogd: Software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

De Bruin A (2006) *The potential for coevolution in aquatic host-parasite system*. Thesis, Radboud University, Nijmegen.

De Bruin A, Ibelings BW, Rijkeboer M, Brehm M, van Donk E (2004) Genetic variation in *Asterionella formosa* (*Bacillariophyceae*): is it linked to frequent epidemics of host-specific parasitic fungi? *Journal of Phycology*, **40**, 823–830.

de Wit R, Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*, **8**, 755–758.

Dolan JR, Ritchie ME, Ras J (2007) The "neutral" community structure of planktonic herbivores, tintinnid ciliates of the microzooplankton, across the SE Tropical Pacific Ocean. *Biogeosciences*, **4**, 297–310.

Dorken ME, Eckert CG (2001) Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology*, **89**, 339–350.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Erdner DL, Richlen M, McCauley LAR, Anderson DM (2011) Diversity and dynamics of a widespread bloom of the toxic dinoflagellate *Alexandrium fundyense*. *PLoS ONE*, **6**, 222965.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Evans KM, Bates SS, Medlin LK, Hayes PK (2004) Microsatellite marker development and genetic variation in the toxic marine diatom *Pseudo-nitzschia multiseries* (Bacillariophyceae). *Journal of Phycology*, **40**, 911–920.

Evans KM, Chepurnov VA, Sluiman HJ *et al.* (2009) Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. *Protist*, **160**, 386–396.

Evans KM, Kuhn SF, Hayes PK (2005) High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. *Journal of Phycology*, **41**, 506–514.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **29**, 342–350.

Fenchel T, Finlay BJ (2004) Here and there or everywhere? Response from Fenchel and Finlay. *BioScience*, **54**, 884–885.

Fraley C (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97**, 611–631.

Fraley C, Raftery A, Murphy T, Scrucca L (2012) *MCLUST Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*, in Technical Report no. 597. Department of Statistics, University of Washington, USA.

Gautschi B, Tenzer I, Muller JP, Schmid B (2000a) Isolation and characterization of microsatellite loci in the bearded vulture (*Gypaetus barbatus*) and cross-amplification in three Old World vulture species. *Molecular Ecology*, **9**, 2193–2195.

Gautschi B, Widmer A, Koella J (2000b) Isolation and characterization of microsatellite loci in the dice snake (*Natrix tessellata*). *Molecular Ecology*, **9**, 2191–2193.

Gsell AS, de Senerpont Domis LN, Przytulska-Bartosiewicz A *et al.* (2012) Genotype-by-temperature interactions may help to maintain clonal diversity in *Asterionella formosa* (Bacillariophyceae). *Journal of Phycology*, **48**, 1197–1208.

Gsell AS, de Senerpont Domis L, Verhoeven K, van Donk E, Ibelings BW (2013) Chytrid epidemics may increase genetic diversity of a diatom spring bloom. *ISME Journal*, **7**, 2057–2059.

Gulati RD, van Donk E (2002) Lakes in the Netherlands, their origin, eutrophication and restoration: state-of-the-art review. *Hydrobiologia*, **478**, 73–106.

Halkett F, Simon JC, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, **20**, 194–201.

Härnström K, Ellegaard M, Andersen TJ, Godhe A (2011) Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4252–4257.

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation G(ST), D(EST) and G'(ST): how wrong have we been? *Molecular Ecology*, **18**, 2080–2083.

Hutchinson GE (1961) The paradox of the plankton. *The American Naturalist*, **95**, 137–145.

Iglesias-Rodriguez MD, Schofield OM, Batley J, Medlin LK, Hayes PK (2006) Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): the use of microsatellite analysis in marine phytoplankton population studies. *Journal of Phycology*, **42**, 526–536.

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

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

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics*, **5**, 539–543.

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**, 625–632.

Karvonen A, Seehausen O (2012) The role of parasitism in adaptive radiations—when might parasites promote and when might they constrain ecological speciation? *International Journal of Ecology*. doi: 10.1155/2012/280169.

Koester JA, Swalwell JE, von Dassow P, Armbrust EV (2010) Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evolutionary Biology*, **10**, 1.

Leibold MA, Norberg J (2004) Biodiversity in metacommunities: plankton as complex adaptive systems? *Limnology and Oceanography*, **49**, 1278–1289.

Lewis RJ, Jensen SI, DeNicola DM *et al.* (1997) Genetic variation in the diatom *Fragilaria capucina* (Fragilariaceae) along a latitudinal gradient across North America. *Plant Systematics and Evolution*, **204**, 99–108.

Lewis WM (1984) The diatom sex clock and its evolutionary significance. *The American Naturalist*, **123**, 73–80.

Logares R, Boltovskoy A, Bensch S, Laybourn-Parry J, Rengefors K (2009) Genetic diversity patterns in five protist species occurring in lakes. *Protist*, **160**, 301–317.

Lowe CD, Montagnes DJS, Martin LE, Watts PC (2010) High genetic diversity and fine-scale spatial structure in the marine flagellate *Oxyrrhis marina* (Dinophyceae) uncovered by microsatellite loci. *PLoS ONE*, **5**, e15557.

Mann DG (1988) Why didn't Lund see sex in *Asterionella*? A discussion of the diatom cycle in nature. In: *Algae and the*

Aquatic Environment (ed. Round FE), pp. 385–412. Biopress Ltd, Bristol, UK.

Mann DG (1999) The species concept in diatoms. *Phycologia*, **38**, 437–495.

Mann DG (2010) Discovering diatom species: is a long history of disagreements about species-level taxonomy now at an end? *Plant Ecology and Evolution*, **143**, 251–264.

Masseret E, Grzebyk D, Nagai S *et al.* (2009) Unexpected genetic diversity among and within populations of the toxic dinoflagellate *Alexandrium catenella* as revealed by nuclear microsatellite markers. *Applied and Environmental Microbiology*, **75**, 2037–2045.

Medlin LK (2007) If everything is everywhere, do they share a common gene pool? *Gene*, **406**, 180–183.

Montero-Pau J, Gomez A, Munoz J (2008) Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography-Methods*, **6**, 218–222.

Nagai S, Nishitani G, Sakamoto S *et al.* (2009) Genetic structuring and transfer of marine dinoflagellate *Cochlodinium polykrikoides* in Japanese and Korean coastal waters revealed by microsatellites. *Molecular Ecology*, **18**, 2337–2352.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

Pante E, Puillandre N, Viricel A *et al.* (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology*, **24**, 525–544.

Pappas J, Stoermer E (2001) *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae) Taxonomic history and quantitative methods as an aid to valve shape differentiation. *Diatom*, **17**, 47–58.

Poulickova A, Vesela J, Neustupa J, Skaloud P (2010) Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kutz. (Bacillariophyceae) and morphologically similar taxa. *Protist*, **161**, 353–369.

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Rousset F (2008) GENEPOP ′007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ruttner F (1937) Limnologische Studien an einigen Seen der Ostalpen (Seen des Salzkammergutes, des Otscherund Hochschwabgebietes). *Archiv fuer Hydrobiologie*, **32**, 167–319.

Rynearson TA, Armbrust EV (2000) DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom *Ditylum brightwellii*. *Limnology and Oceanography*, **45**, 1329–1340.

Rynearson TA, Armbrust EV (2004) Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology*, **40**, 34–43.

Rynearson TA, Armbrust EV (2005) Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. *Molecular Ecology*, **14**, 1631–1640.

Rynearson TA, Lin EO, Armbrust EV (2009) Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist*, **160**, 111–121.

Rynearson TA, Newton JA, Armbrust EV (2006) Spring bloom development, genetic variation, and population succession in the planktonic diatom *Ditylum brightwellii*. *Limnology and Oceanography*, **51**, 1249–1261.

Santos HM, Paiva M-R, Rocha S, Kerdelhué C, Branco M (2013) Phenotypic divergence in reproductive traits of a moth population experiencing a phenological shift. *Ecology and Evolution*, **3**, 5098–5108.

Scrucca L (2010) Dimension reduction for model-based clustering. *Statistics and Computing*, **20**, 471–484.

Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Smayda TJ (2011) Cryptic planktonic diatom challenges phytoplankton ecologists. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4269–4270.

Sonstebo JH, Rohrlack T (2011) Possible implications of chytrid parasitism for population subdivision in freshwater cyanobacteria of the genus *Planktothrix*. *Applied and Environmental Microbiology*, **77**, 1344–1351.

Soudek D, Robinson GGC (1983) Electrophoretic analysis of the species and population- structure of the diatom *Asterionella formosa*. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **61**, 418–433.

Stein J (1973) *Culture Methods and Growth Measurements. Handbook of Phycological Methods*. Cambridge University Press, Cambridge, UK.

Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. *Molecular Ecology Notes*, **3**, 329–331.

Trobajo R, Clavero E, Chepurnov VA *et al.* (2009) Morphological, genetic and mating diversity within the widespread bioindicator *Nitzschia palea* (Bacillariophyceae). *Phycologia*, **48**, 443–459.

Van den Wyngaert S, Vanholsbeeck O, Spaak P, Ibelings BW (2014) Parasite fitness traits under environmental variation: disentangling the roles of a chytrid's immediate host and external environment. *Microbial Ecology*, **68**, 645–656.

Vanelslander B, Créach V, Vanormelingen P *et al.* (2009) Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom *Navicula phyllepta* (Bacillariophyceae). *Journal of Phycology*, **45**, 1278–1289.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vanormelingen P, Cottenie K, Michels E *et al.* (2008) The relative importance of dispersal and local processes in structuring phytoplankton communities in a set of highly interconnected ponds. *Freshwater Biology*, **53**, 2170–2183.

Vanoverbeke J, De Gelas K, De Meester L (2007) Habitat size and the genetic structure of a cyclical parthenogen, *Daphnia magna*. *Heredity*, **98**, 419–426.

Walser B, Haag CR (2012) Strong intraspecific variation in genetic diversity and genetic differentiation in *Daphnia magna*: the effects of population turnover and population size. *Molecular Ecology*, **21**, 851–861.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.

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

Zhang W-J, Yang J, Yu Y-H, Shu S-W, Shen Y-F (2006) Population genetic structure of *Carchesium polypinum* (Ciliophora: Peritrichia) in four Chinese lakes inferred from ISSR fingerprinting: high diversity but low differentiation. *Journal of Eukaryotic Microbiology*, **53**, 358–363.

S.V.D.W. performed the field and laboratory work. S.V.D.W., M.M. and R.M. analysed the data. S.V.D.W. and M.M. wrote the manuscript. P.S. and B.W.I. supervised the project. All authors critically revised and approved the final manuscript.

Data accessibility

The following datafiles are archived in the Dryad Digital Repository: doi:10.5061/dryad.3k0d8.

Datafile containing microsatellite genotype data of *Asterionella formosa* isolates from seven Swiss lakes and four Dutch lakes.

Datafile containing AFLP data of a subset of isolates of *A. formosa*.

Datafile containing the microsatellite sequences of *A. formosa*.

Supporting information

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

Fig. S1 The steady decrease of BIC values shows that the inference of the number of clusters in the DAPC analysis is not straightforward.

Fig. S2 Clustering of all unique MLGs of *Asterionella formosa* by Structure, represented for different number of *K*'s ranging from $K = 2$ until $K = 9$.

Fig. S3 Box plot of frustules cell length of upper and Lower Lake Zurich (ZSU and ZSO), Lake Baldegg (BS), Lake Greifen late sampling (GSL), Lake Lucerne basin B (VWSB) and Lake Halwill (HW) within each of the clusters A and B.

Table S1 Number of alleles observed on each locus in each population (N_a), sample sizes (N), expected (H_e) and observed (H_o) heterozygosities, F_{IS}, probabilities of the Hardy–Weinberg test for a heterozygote deficiency (P-val), standard error of F_{IS} from a locus to the other in each population (SE).

Table S2 P_{sex} values and their significance calculated with MLGSIM2.0.