FEMS Microbiology Ecology, 93, 2017, fix125

doi: 10.1093/femsec/fix125
Advance Access Publication Date: 3 October 2017
Research Article

RESEARCH ARTICLE

Geographical distance and local environmental conditions drive the genetic population structure of a freshwater microalga (Bathycoccaceae; Chlorophyta) in Patagonian lakes

Leonardo D. Fernández[1,2,3,*,†], Cristián E. Hernández[3], M. Romina Schiaffino[4], Irina Izaguirre[5] and Enrique Lara[2,6]

[1]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile, [2]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland, [3]Laboratorio de Ecología Evolutiva y Filoinformática, Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Barrio Universitario s/n, 4030000 Concepción, Chile, [4]Centro de Investigaciones y transferencia del Noroeste de la Provincia de Buenos Aires, CONICET, UNNOBA, Junín 6000, Argentina, [5]Departamento de Ecología, Genética y Evolución, IEGEBA (CONICET-UBA), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Buenos Aires C1428EHA, Argentina and [6]Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain

[*]**Corresponding author:** Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile. Tel: +56-224772228; E-mail: limnoleo@gmail.com

One sentence summary: The unexplained variation in the genetic population structure of a freshwater Bathycoccaceae microalga recently discovered in lakes of South America could be the result of founder events combined with rapid local adaptations, as proposed by the monopolisation hypothesis.

Editor: Hendrikus Laanbroek

[†]Leonardo D. Fernández, http://orcid.org/0000-0001-9550-1921

ABSTRACT

The patterns and mechanisms underlying the genetic structure of microbial populations remain unresolved. Herein we investigated the role played by two non-mutually exclusive models (i.e. isolation by distance and isolation by environment) in shaping the genetic structure of lacustrine populations of a microalga (a freshwater Bathycoccaceae) in the Argentinean Patagonia. To our knowledge, this was the first study to investigate the genetic population structure in a South American microorganism. Population-level analyses based on ITS1–5.8S-ITS2 sequences revealed high levels of nucleotide and haplotype diversity within and among populations. Fixation index and a spatially explicit Bayesian analysis confirmed the occurrence of genetically distinct microalga populations in Patagonia. Isolation by distance and isolation by environment accounted for 38.5% and 17.7% of the genetic structure observed, respectively, whereas together these models accounted for 41% of the genetic differentiation. While our results highlighted isolation by distance and isolation by environment as important mechanisms in driving the genetic population structure of the microalga studied, none of these models (either

Received: 20 June 2017; **Accepted:** 26 September 2017

alone or together) could explain the entire genetic differentiation observed. The unexplained variation in the genetic differentiation observed could be the result of founder events combined with rapid local adaptations, as proposed by the monopolisation hypothesis.

Keywords: freshwater protist; gene flow; Mamiellophyceae; microbial populations; phytoplankton; picoeukaryote

INTRODUCTION

Free-living aquatic microorganisms are among the most diverse, ubiquitous and abundant living entities on Earth (Logares *et al.* 2014; de Vargas *et al.* 2015). Because of their huge populations, small size and capability to form dormant stages to survive adverse environmental conditions, microorganisms were largely assumed to display unlimited passive dispersal (Beijerinck 1913; Finlay 2002; Fenchel and Finlay 2004) and, accordingly, unrestricted gene flow among populations (Martiny *et al.* 2006; Rengefors, Logares and Laybourn-Parry 2012). While some highly abundant and euryoecious marine microorganisms appear to have unlimited dispersal and a lack of genetic structure among populations (Evans, Kühn and Hayes 2005; Šlapeta, Lopez-Garcia and Moreira 2006; Lara *et al.* 2009; Casteleyn *et al.* 2010; Van Gremberghe *et al.* 2011), a growing body of evidence supports the idea that many microorganisms do have limited dispersal and population genetic structure. Broad-scale dispersal of water-dependent protists, for example, is limited by present-day processes, including environmental factors (Vyverman *et al.* 2007; Fernández *et al.* 2016), geographical barriers (Fournier *et al.* 2016) or both (Schiaffino *et al.* 2016). Even, there is evidence showing that in protists, dispersal capability can be restricted by long-term processes, such as evolutionary constraints to water deficit and/or extreme temperatures (Fernández *et al.* 2016). On the other hand, several marine microorganisms exhibit genetic differentiation among neighbouring populations despite dispersal potential through currents (Rynearson, Newton and Armbrust 2006; Nagai *et al.* 2007; Godhe and Härnström 2010; Härnström *et al.* 2011; Sjöqvist *et al.* 2015). Microorganisms inhabiting highly connected inland waters have also revealed complex population genetic structure comparable with macroorganisms (Kim *et al.* 2004; Evans *et al.* 2009; Logares *et al.* 2009; Van den Wyngaert *et al.* 2015). Limited gene flow and genetic structure has even been reported among lacustrine populations of microorganisms that were separated by less than 10 km (Rengefors, Logares and Laybourn-Parry 2012).

Recent research suggests that the genetic structure of microbial populations may be driven by two non-mutually exclusive models, namely isolation by distance and isolation by environment (Van Gremberghe *et al.* 2011; Rengefors, Logares and Laybourn-Parry 2012; Van den Wyngaert *et al.* 2015). Isolation by distance (Wright 1943) proposes that geographical distance and barriers restrict gene flow among populations, leading to a positive correlation between genetic and geographical distances. Isolation by environment (Wang and Summers 2010) proposes that local environmental conditions select against migrants, leading to limited gene flow among populations and so, to a positive correlation between genetic and environmental distances.

Isolation by distance has been invoked to explain the role that currents and oceanic basins might be playing in limiting dispersal and gene flow among marine microbial populations (Casabianca *et al.* 2012; Godhe *et al.* 2013), or the role that the lack of hydrological connectivity and long-term ice cover play in restricting dispersal and gene flow among lacustrine microbial populations (Rengefors, Logares and Laybourn-Parry 2012). Isolation by environment, on the other hand, appears to have a large impact on the genetic structure of microbial populations inhabiting different but highly connected marine (de Vargas *et al.* 2002; Rodríguez-Martínez *et al.* 2013) and estuarine (Rynearson, Newton and Armbrust 2006; Härnström *et al.* 2011) environments. However, the role that this model might play in shaping the genetic structure of microbial populations in disconnected water systems, such as in lakes, remains so far untested.

Most of the studies that have assessed the patterns and causes of spatial genetic diversity and structure in microbial populations have been focused on aquatic ecosystems of the northern hemisphere (e.g. Rynearson, Newton and Armbrust 2006; Godhe and Härnström 2010; Härnström *et al.* 2011; Van Gremberghe *et al.* 2011; Casabianca *et al.* 2012; Lowe *et al.* 2012; Godhe *et al.* 2013; Van den Wyngaert *et al.* 2015); to our knowledge, there are currently no similar studies on microorganisms from historically undersampled areas, such as Patagonia in southern South America.

Patagonia is regarded as an endemism hotspot for protists (Vucetich and Lopretto 1995; Woelfl 2006; Fernández, Lara and Mitchell 2015) and exhibits countless lakes that represent an interesting scenario to investigate the patterns and mechanisms underlying the genetic population structure of microorganisms. These lakes span extensive latitudinal and environmental gradients, are hydrologically disconnected from each other and have a patchy distribution in Patagonia (Quirós and Drago 1999; Tell, Izaguirre and Allende 2011; Izaguirre *et al.* 2015). Therefore, microorganisms living in Patagonian lakes probably exhibit high levels of genetic diversity and genetic population structure because both geographical distance between lakes (isolation by distance) and ecological conditions of each lake (isolation by environment) would synergistically limit the gene flow among their populations.

Herein we tested this hypothesis on lake populations of a new and still undescribed Mamiellophyceae microalga, which was recently revealed by a molecular survey conducted on a set of lakes situated in the Argentinean Patagonia (Schiaffino *et al.* 2016). This microalga was later found to be phylogenetically related to *Bathycoccus prasinos*, thus becoming the only known freshwater member of family Bathycoccaceae (Lara *et al.* 2017). This newly erected chlorophyte family (Marin and Melkonian 2010) comprises the smallest known photosynthetic eukaryotes (i.e. picoeukaryotes), *B. prasinos* and *Ostreococcus tauri* (Courties *et al.* 1994; Marin and Melkonian 2010). The extreme abundance of these organisms in marine environments (Countway and Caron 2006), coupled with their small size, suggests that their geographical distributions should tend toward cosmopolitanism (Fenchel and Finlay 2004). Indeed, such trends have been observed, and genotypes are better separated by different environmental conditions than by geographical distance in marine members of Bathycoccaceae (Demir-Hilton *et al.* 2011). However, as hypothesised above, freshwater environments should promote higher population

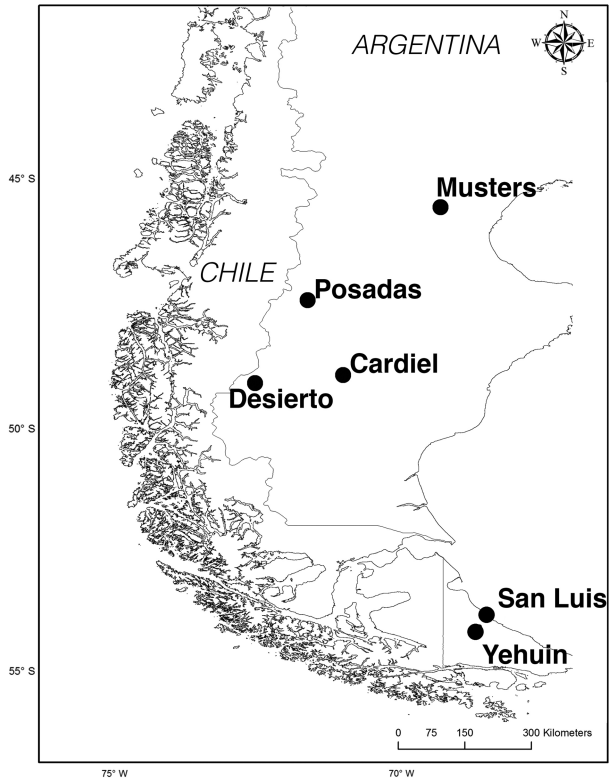

Figure 1. Map of the study site showing the geographical distribution of the six lakes studied in the Argentinean Patagonia, South America.

genetic structure in microorganisms via isolation by distance and isolation by environment.

MATERIALS AND METHODS

Sampling strategy

We collected water samples in the euphotic zone of 40 lakes distributed from Chubut province (Patagonia, Argentina) to Hope Bay (Antarctic Peninsula). Antarctic lakes were sampled during the austral summer of 2004, whereas Patagonian lakes were sampled in late spring 2007 and 2008. At each lake, we measured *in situ* environmental variables, including temperature, pH, conductivity and dissolved oxygen using a portable probe (Horiba D-54 meter, Kyoto, Japan and Hanna HI 9146, Villafranca, Italy). Once back in the laboratory, we also measured supplementary environmental variables, including ammonium, chlorophyll a, dissolved organic carbon, nitrate, nitrite and soluble reactive phosphorus following methods described in Schiaffino *et al.* (2011). Additional information on the studied lakes can be found in Tell, Izaguirre and Allende (2011) and Izaguirre *et al.* (2015), whereas the main environmental features of the Argentinean Patagonian limno-regions are given in Quirós and Drago (1999).

A subsequent fingerprinting-based survey of these water samples (Schiaffino *et al.* 2016) revealed that 6 of the 40 lakes studied harboured a previously unknown freshwater member of family Bathycoccaceae (Lara *et al.* 2017). These six lakes were in the Argentinean Patagonia, specifically in the provinces of Chubut (Musters Lake), Santa Cruz (Posadas, Cardiel and Desierto Lakes) and Tierra del Fuego (San Luis and Yehuin Lakes) (Fig. 1, Table 1). All molecular and genetic population structure analyses described below were therefore exclusively conducted on samples from these six lakes.

Molecular methods

Water samples for nucleic acid extractions were pre-filtered *in situ* through a 50 μm net and about 1 l of water of each lake was filtered sequentially through 20, 3 and 0.2 μm pore size polycarbonate filters with a vacuum pump. The filters were preserved in cryovials with 1.8 mL of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) and stored at –80°C until nucleic acid extraction. The 0.2–3 μm size fraction was used for this study. Nucleic acid extractions were performed with a phenol/chloroform procedure as described in Unrein *et al.* (2005) and extracts were stored at –80°C until analysis.

We used an alignment comprising SSU and LSU rRNA sequences of several representatives of the Bathycoccaceae/Mamiellophyceae clade, both environmental and derived from cultures (cloning). We chose the primers to amplify a region that could encompass the whole ITS operon (i.e. the ITS1-5.8S-ITS2 complex), and to be specific only to this clade and not to amplify Monomastigales, which can be common in freshwaters (Marin and Melkonian 2010). The fragment that was sequenced included a part of the 18S gene, which included the v9 hypervariable region, one of the most frequently used markers in environmental eukaryotic DNA-based diversity surveys. This region was invariable in all clones, thus demonstrating that each of them belonged to the novel lineage of freshwater Bathycoccaceae. The resulting primers, PCR procedure, cloning/sequencing protocols, as well as the identity and phylogenetic position of the new freshwater Bathycoccaceae were fully provided in Lara *et al.* (2017).

Population data analyses

Descriptive statistics of genetic diversity, including the number of haplotypes (*h*), haplotype diversity (*Hd*), nucleotide diversity (π) and number of segregating sites, were estimated for each population using DnaSP version 5.10.1 (Librado and Rozas 2009).

Table 1. Molecular diversity for the ITS region at each sampled lake in the Argentinean Patagonia.

Lake	Latitude (°S)	Longitude (°W)	Sampling	N	h	S	Hd (±SD)	π
Musters	45°33′00″	69°08′24″	Spring 2007	36	4	5	0.610 ± 0.05	0.00160
Posadas	47°27′00″	71°49′12″	Spring 2007	27	3	2	0.211 ± 0.10	0.00037
Cardiel	48°59′24″	71°07′48″	Spring 2007	31	3	3	0.411 ± 0.09	0.00195
Desierto	49°04′47″	72°53′01″	Spring 2007	25	4	7	0.757 ± 0.04	0.00539
San Luis	53°55′19″	67°36′59″	Spring 2008	29	3	3	0.493 ± 0.09	0.00158
Yehuin	54°21′39″	67°46′49″	Spring 2008	20	8	10	0.879 ± 0.05	0.00527
All	–	–	–	168	18	20	0.787 ± 0.03	0.00364

N = sample size; h = number of haplotypes; S = number of segregating sites; Hd = haplotype diversity; π = nucleotide diversity. Lakes are sorted from north to south.

A haplotype network was constructed using a Median-Joining algorithm implemented in the program NETWORK version 4.5 (http://www.fluxus-technology.com) (Bandelt, Forster and Röhl 1999) to determine the genealogical relationship between the haplotypes and the spatial distribution of the haplotypes. Genetic divergence among populations was estimated by calculating the pairwise genetic differentiation (pairwise F_{ST} statistics) in ARLEQUIN version 3.5.2.2 (Excoffier and Lischer 2010) based on 50 000 permutations ($P < 0.05$). Sequential Bonferroni corrections were used to adjust significance levels for simultaneous inferential tests (Rice 1989).

Genetic population structure

The existence of spatial discontinuities (i.e. the existence of genetic structure) among populations was investigated using the program Geneland version 4.0.5 (Guillot *et al.* 2005; Guillot, Mortier and Estoup 2005; Guillot, Santos and Estoup 2008). Geneland implements a population statistical model with Bayesian inference in a set of georeferenced individuals with sequences data. The goal of this model is to infer and locate the genetic discontinuities between populations of georeferenced genotypes, considering the uncertain localisation of the sampled individuals. The number of clusters was determined by running five Markov Chain Monte Carlo (MCMC) runs and allowing K (i.e. the most probable number of populations) to vary according to the following parameters: 10 000 000 MCMC iterations, maximum rate of the Poisson process fixed to 100 (this is the default value of the software), the minimum K = 1, maximum K = 6 (values that allow exploring a wide potential number of populations, and considering the maximum spatial subdivision in the latitudinal gradient considered). The maximum number of nuclei in the Poisson-Voronoi tessellation was fixed at 300 (3 × maximum rate was suggested by Guillot *et al.* 2005; Guillot, Mortier and Estoup 2005). After inferring the number of populations (clusters) in the data set from these five MCMC runs, we re-run two new MCMC runs, with K fixed to the inferred number of clusters and leaving the other parameters the same as above. The posterior probability of population membership for each pixel of the spatial domain was then computed.

Isolation by distance and isolation by environment

We used multiple regression on distance matrices (MRM; Legendre and Legendre 1998; Lichstein 2007) of genetic, geographical and environmental information to investigate the extent to which isolation by distance and isolation by environment (either together or alone) influence the genetic connectivity and structure of microbial populations. Unlike other methods for analysing correlations between distance matrices (e.g. Mantel test; Mantel 1967), MRM performs a multiple regression analysis between two or more distance matrices, using permutations to determine the significance of the coefficients of determination (Legendre and Legendre 1998; Lichstein 2007). MRM was conducted using Slatkin's linearised F_{ST} as response and geographical and environmental distance matrices as predictors. The geographical distance matrix was calculated with the geographical distances between each pair of lakes, using their geographical coordinates and Euclidean distances. To construct the environmental distance matrix, we log transformed environmental variables to avoid skewed trends and then checked for multicollinearity. Next, we normalised all variables to standardised the different units of the different variables (Clarke *et al.* 2005), and used these values to build up a distance matrix based on Euclidean distances. MRM analyses were conducted using 50 000 random permutations of the rows and columns of the dependent matrix (while keeping the other matrices fixed) in the package ecodist (Goslee and Urban 2007) implemented in R version 3.2.3 (R Core Team 2013). Sequential Bonferroni corrections were used to adjust significance levels for simultaneous inferential tests (Rice 1989).

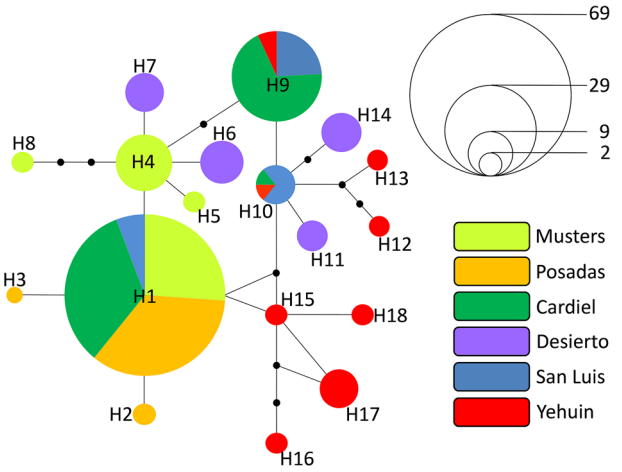

Figure 2. Haplotype network of the Argentinean freshwater Bathycoccaceae based on internal transcribed spacer operon sequences. Each haplotype has been identified with a letter–number combination and resized according to its frequency. Black dots on connecting lines indicate the number of mutational steps between haplotypes. The colours making up each haplotype represent the lacustrine origin of the sequences that compose it.

RESULTS

We obtained positive PCR products for six lakes located in the Argentinean Patagonia (Musters, Posadas, Cardiel, Desierto, San Luis and Yehuin) out of a gradient of 40 lakes spanning from Patagonia to the Antarctica (Schiaffino *et al.* 2016). We cloned the PCR products resulting in a total of 168 ITS sequences (593 bp). For the whole study site, genetic diversity indices revealed a total of 18 haplotypes (h) and 20 segregating sites (S), as well as a total haplotype (Hd) and nucleotide (π) diversity of 0.787 and 0.00364, respectively (Table 1). Genetic diversity indices varied widely among lake populations. Excepted for nucleotide diversity, the southernmost lake population (Yehuin) exhibited the highest values of genetic diversity indices, followed by the northernmost (Musters) and central (Desierto) lake populations, respectively (Table 1). The three other populations (Posadas, Cardiel and San Luis) exhibited comparatively lower values for these indices (Table 1). Overall, the number of haplotypes ranged from 8 to 3, the number of segregating sites between 10 and 2, haplotype diversity between 0.879 and 0.211, and nucleotide diversity between 0.00539 and 0.00037 (Table 1).

The haplotype network suggested that haplotypes have a geographically structured distribution in the Argentinean Patagonia (Fig. 2). Haplotypes H1 and H9 were the most common in Patagonia, although none of them was present in all lakes. Several other haplotypes were only present at specific lakes. Particularly, noteworthy were those present in the central lake Desierto and the southernmost lake Yehuin. Four and six haplotypes were exclusive to these lakes, respectively.

All pairwise F_{ST} values were significant after Bonferroni correction for multiple testing ($\alpha = 0.05$). Genetic divergence between all lake populations was moderate to high, ranging

Table 2. Estimates of F_{ST} (below diagonal) and corresponding P values (above diagonal) based on the ITS variation between populations of the freshwater Bathycoccaceae over the Argentinean Patagonia.

	Musters	Posadas	Cardiel	Desierto	San Luis	Yehuin
Musters	–	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Posadas	0.285	–	<0.0001	<0.0001	<0.0001	<0.0001
Cardiel	0.154	0.191	–	<0.0001	<0.0001	<0.0001
Desierto	0.264	0.376	0.173	–	<0.0001	<0.0001
San Luis	0.615	0.771	0.465	0.240	–	<0.0001
Yehuin	0.434	0.452	0.281	0.256	0.397	–

All F_{ST} values were significant after sequential Bonferroni correction for multiple testing ($\alpha = 0.05$). The diagonal values are zero. Lakes are sorted from north to south.

between 0.154 and 0.771 (Table 2). Even geographically close lacustrine populations (Yehuin and San Luis, which are just ca. 50 km apart) exhibited high genetic divergence (Table 2). The highest genetic divergence was found between the populations of lakes Posadas and San Luis: two demes that are in areas geographically wide apart (ca. 770 km) (Table 2).

The Geneland analysis inferred three genetic clusters distributed across the northern, central and southern Patagonia (Fig. 3). The first cluster was composed of populations occurring in lakes Musters, Posadas and Cardiel. The second cluster was composed of populations occurring in lakes Desierto and San Luis. The third cluster only included the population of the southernmost lake, Yehuin.

The MRM analysis showed that isolation by distance explained a higher proportion of the genetic variation among populations than isolation by environment (Table 3). However, this analysis also revealed that together both models accounted for a higher proportion of the genetic structured observed than when they are considered alone (Table 3).

DISCUSSION

This study supported the hypothesis that lacustrine microbial populations exhibit high levels of genetic structure driven by isolation by distance and isolation by environment.

The populations studied exhibited moderate to high levels of genetic divergence, with F_{ST} values ranging between 0.154 and 0.771 (Table 2). While our results are not directly comparable to those based on microsatellite data, it was evident that the F_{ST} values observed in Patagonian lakes were higher than those reported for other genetically structured populations of aquatic eukaryotic microorganisms, such as lake-dwelling dinoflagellates (F_{ST} values, 0.117–0.376; Rengefors, Logares and Laybourn-Parry 2012) and brackish water diatoms (F_{ST} values, 0.109–0.286; Rynearson, Newton and Armbrust 2006). The high F_{ST} values observed in our study could only be explained in terms of limited dispersal and gene flow among populations. This conclusion was supported by a spatially explicit Bayesian clustering approach (Geneland; Fig. 3), which revealed that gene flow is particularly limited among lacustrine populations located to the north, central and south of Patagonia.

Isolation by distance and isolation by environment

In agreement with other studies on genetic population structure (Ogden and Thorpe 2002; Crispo *et al.* 2006; Thorpe, Surget-Groba and Johansson 2008; Lee-Yaw *et al.* 2009; Lee and Mitchell-Olds 2011), the MRM analysis revealed that a high proportion of the genetic population structure observed resulted from the joint influence of isolation by distance and isolation by environment (Table 3). A recent simulation analysis showed that the use of MRM is appropriate to quantify the effects of isolation by distance and isolation by environment when dispersal rates are low or moderate (Wang 2013). Our analyses of genetic population structure (F_{ST} and Geneland analyses) suggested that our model microorganism exhibits low to moderate dispersal and gene flow, and therefore we considered our MRM output robust enough to support our conclusions.

The model organism of this study is included within Bathycoccaceae (Lara *et al.* 2017), a family comprising very small microalgae with cells ranging from 0.5 to 2.5 μm in size (Marin and Melkonian 2010). This freshwater Bathycoccaceae inhabits lakes that are influenced by a variety of natural and human dispersal vectors, including migratory water birds (Paruelo *et al.* 2005) and sport fishermen (Quirós and Drago 1999), that could potentially contribute to passively disperse their propagules towards distant freshwater bodies. Simulation analyses (Wilkinson *et al.* 2012) and empirical evidence (Wilkinson 2001; Yang *et al.* 2010; Lara *et al.* 2011) also suggest that smaller microorganisms can be easily airborne dispersed over large spatial scales. Therefore, from a theoretical point of view, this freshwater Bathycoccaceae should not have any population genetic structure as this would potentially be erased by fast dispersal (Finlay 2002; Fenchel and Finlay 2004) and gene flow (Martiny *et al.* 2006; Rengefors, Logares and Laybourn-Parry 2012). In spite of this, our analyses supported isolation by distance as an important contributor to patterns of spatial genetic variation in populations of our model microorganism. This contrasts with observations on marine picoeukaryotes, such as member of genera *Ostreococcus*, *Bathycoccus* (Demir-Hilton *et al.* 2011; Simmons *et al.* 2016) and stramenopiles (Rodríguez-Martínez *et al.* 2013), where intraspecific diversity seems to be driven mainly by local environmental conditions (i.e. isolation by environment) than by geographical distance among populations.

Possibly, geographical isolation is an important contributor to patterns of genetic population structure in our model organism because it does not have the appropriate adaptations to travel over large distances as an active cell or under a cystic stage. Active cells of aquatic eukaryotic microorganisms are potentially vulnerable to dehydration during overland transport (Weisse 2008), and we do not know if this microalga is capable of forming a resting stage (e.g. a cyst) that could allow individuals to survive and effectively disperse to other lakes under adverse environmental conditions. *Mantoniella*, a genus contained within class Mamiellophyceace, can form cysts (Ellegaard *et al.* 2016), suggesting that this could also be the case for our model organisms and other microalgae included within Bathycoccaceae. Notwithstanding this, the viability of cysts varies widely between and within species (Foissner 2007), and all of them lack adaptations for air dispersal (Foissner 2008). Therefore, not all cysts necessarily survive under adverse conditions. This might explain why isolation by distance frequently appears as an

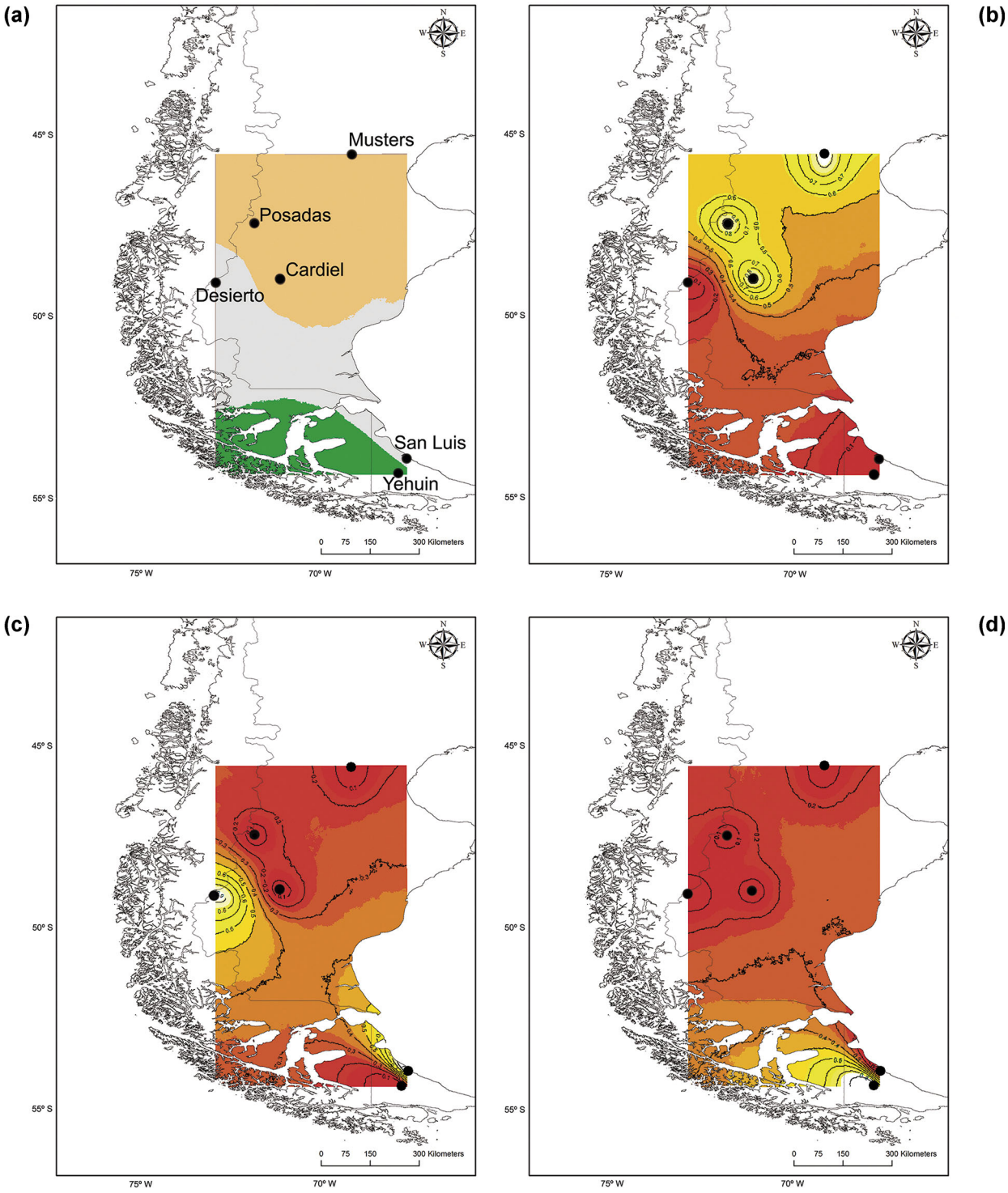

Figure 3. Spatial Bayesian clustering in the program Geneland. Geneland output has been cropped, re-scaled and superimposed on the map of the Argentinean Patagonia. Black dots depict the studied lakes. Lighter colours indicate higher probabilities of population membership. Contour lines indicate the spatial position of genetic discontinuities between populations. (**A**) Spatial clustering suggests three distinct clusters across the Argentinean Patagonia. Map of posterior probabilities of population membership and spatial location of genetic discontinuities for populations of the freshwater Bathycoccaceae microalga belonging to (**B**) the north-central cluster (**C**) the south-central cluster and (**D**) the southernmost cluster.

Table 3. Genetic divergence between populations explained by geographic and environmental distances.

	Response distance matrix	
Explanatory distance matrix	R^2	P
Geographic + environmental distance	0.410	<0.0001
Geographic distance	0.385	<0.0001
Environmental distance	0.177	<0.0001

All values were significant after Bonferroni correction for multiple testing ($\alpha = 0.05$).

important contributor to patterns of spatial genetic variation in populations of aquatic microorganisms (Casteleyn *et al.* 2010; Rengefors, Logares and Laybourn-Parry 2012; Demura *et al.* 2014).

Moreover, our analyses demonstrated that environmental isolation also contributed to explain part of the differentiation among lake populations. Patagonian lakes differ significantly in terms of their abiotic characteristics. Lake Posadas and Yehuin are oligotrophic, lake Cardiel is oligo-mesotrophic, lake Muster is mesotrophic and lake San Luis is eutrophic (Quirós and Drago 1999; Tell, Izaguirre and Allende 2011). Besides their particular physicochemical conditions (Schiaffino *et al.* 2011), these lakes are also distributed over an extensive geographical area, and thus, they are subjected to different climates (Paruelo *et al.* 1998). Consequently, Patagonian lakes represent environmentally contrasting islands surrounded by a suboptimal soil matrix that may be potentially limiting the dispersal and gene flow among populations inhabiting different lakes via selection against dispersers (Nosil 2004; Nosil, Vines and Funk 2005). In agreement with a few previous studies (de Vargas *et al.* 2002; Rynearson, Newton and Armbrust 2006; Härnström *et al.* 2011), we conclude that isolation by environment also have an important, although frequently overlooked, impact on the genetic structuring of microbial populations. This should, however, be further tested over larger scales and using more sites and ideally more taxonomic groups.

A complementary cause for the observed genetic population structure

The genetic differentiation observed between lake populations could also be complementary explained in terms of the monopolisation hypothesis (De Meester *et al.* 2002). This hypothesis states that genetic differentiation between populations can be explained by rapid population growth and local adaptation after historical founder effect, which result in the effective monopolisation of resources (a priority effect). Once a population is locally adapted, the presence of a large resting propagule bank provides a buffer against new immigrants, enhancing priority effects. This hypothesis has been already used to explain population genetic structure in lake-dwelling microorganisms (Rengefors, Logares and Laybourn-Parry 2012).

Northern lakes in Patagonia (Musters, Posadas, Cardiel) are much older than those located in the centre (Desierto) and the south (San Luis and Yehuin). Northern lakes originated during the Mesozoic through tectonic processes (Vilela 1971; Bellosi 1995; González Díaz and Di Tommaso 2014), while lakes of central and southern Patagonia originated during the Holocene after the retreat of the ice sheet at the end of the Last Glacial Maximum (Vilela 1971; Mariazzi *et al.* 1987). These antecedents and our data suggest that populations in central and southern lakes descend from ancestral colonisers coming from northern lakes.

Overall, haplotype diversity was high but nucleotide diversity was low (Table 1), suggesting recent differentiation with rapid population growth in the Argentinean Patagonia (Grant and Bowen 1998; Muller, Leppälä and Savolainen 2008; Palma *et al.* 2012). Also, the haplotype network (Fig. 2) revealed a few common haplotypes intermixed with several unique haplotypes mainly restricted to the northwestern and southern lakes. Haplotype H1 was the most widespread and abundant haplotype in the study site, and therefore, could likely be the ancestral haplotype of this microalga. Besides having many closely related derivates and a more or less central position in the network, this haplotype was also represented by sequences coming from the oldest, north central lakes. Geneland analyses also identified north central lake populations as being genetically distinct from those present in the northwest and the south, supporting a founder effect induced by postglacial colonisation in the latter areas. Furthermore, these analyses identified the populations of Lakes Desierto (northwestern Patagonia) and San Luis (southern Patagonia) as a group genetically distinct from the population of Lake Yehuin (the southernmost lake), suggesting that the founder populations of these lakes were different. Probably, these founder populations followed rapid population growth and local adaptation to these younger lakes. Therefore, as suggested by the monopolisation hypothesis, the observed genetic differentiation between populations may also be due to biological barriers. Individuals may disperse well between lakes, but be unable to colonise because they cannot outcompete residents.

CONCLUSIONS

We conclude that lacustrine systems promote high levels of genetic structure in microbial populations mainly via isolation by distance and to a lesser extent via isolation by environment, although we do not rule out the possible role played by other processes (e.g. biotic, historical). To our knowledge, this study was the first to investigate the patterns and causes of genetic population structure in a free-living microorganism from South America. Therefore, our study not only contributed to disentangle the underlying causes that drive the genetic population structure in microorganisms but also opened a new research line in this continent. Considering that this study was conducted in a remote and historically undersampled region, it is our hope that our results will contribute to fill gaps in knowledge and to construct generalisations in microorganisms from both the northern and southern hemispheres.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGEMENT

We would like to express our gratitude to Amandine Pillonel for her help at the molecular biology laboratory and colleagues Fernando Unrein, Patricia Rodríguez, Rodrigo Sinistro, M. Laura Sánchez, Guillermo Tell, Adrián Rua for their help with sampling. We also thank Edward Mitchell for constructive criticism and valuables comments.

FUNDING

This study was partially funded by grants of the FonCyT–Argentina (PICT 32732 and 0794) and CONICET–Argentina (PIP 498). LDF was supported by CONICYT–Chile (doctoral grants N° 21110037 and 78130011); Universidad Bernardo O'Higgins (Proyecto interno VRIP N° 170201); Programa de Doctorado en Sistemática y Biodiversidad y la Dirección de Postgrado de la Universidad de Concepción (mobility grant); and the University of Neuchâtel (mobility grant, IDPOB). EL was supported by Swiss National Fund grant SNF 31003A_163254.

Conflict of interest. None declared.

REFERENCES

Bandelt HJ, Forster P, Röhl A. Median–joining networks for inferring intraspecific phylogenies. *Mol Biol Ecol* 1999;**16**:37–48.

Beijerinck MW. *De infusies en de ontdekking der backteriën*. Müller, Amsterdam: Jaarboek van de Koninklijke Akademie van Wetenschappen, 1913.

Bellosi ES. Paleogeografía y cambios ambientales de la Patagonia central durante el Terciario medio. *Bol Inf Pet* 1995;**44**:50–83.

Casabianca S, Penna A, Pecchioli E *et al*. Population genetic structure and connectivity of the harmful dinoflagellate *Alexandrium minutumin* the Mediterranean Sea. *Proc R Soc Lond* 2012;**279**:129–38.

Casteleyn G, Leliaert F, Backeljau T *et al*. Limits to gene flow in a cosmopolitan marine planktonic diatom. *P Natl Acad Sci USA* 2010;**107**:12952–7.

Clarke KR, Warwick RM, SomerWeld PJ *et al*. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 3rd edn. Plymouth, UK: PRIMER-E Ltd, 2005.

Countway PD, Caron DA. Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microb* 2006;**72**:2496–506.

Courties C, Vaquer A, Troussellier M *et al*. Smallest eukaryotic organism. *Nature* 1994;**370**:255.

Crispo E, Bentzen P, Reznick DN *et al*. The relative influence of natural selection and geography on gene flow in guppies. *Mol Ecol* 2006;**15**:49–62.

De Meester L, Gómez A, Okamura B *et al*. The monopolization hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecol* 2002;**23**:121–35.

Demir-Hilton E, Sudek S, Cuvelier ML *et al*. Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *ISME J* 2011;**5**:1095–107.

Demura M, Nakayama T, Kasai F *et al*. Genetic structure of Japanese *Chattonella marina* (Raphidophyceae) populations revealed using microsatellite markers. *Phycol Res* 2014;**62**:102–8.

de Vargas C, Audic S, Henry N *et al*. Eukaryotic plankton diversity in the sunlit ocean. *Science* 2015;**348**:1261605.

de Vargas C, Bonzon M, Rees NW *et al*. A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Mar Micropaleontol* 2002;**45**:101–16.

Ellegaard M, Moestrup O, Andersen TJ *et al*. Long-term survival of haptophyte and prasinophyte resting stages in marine sediment. *Eur J Phycol* 2016;**51**:328–37.

Evans KM, Chepurnov VA, Sluiman HJ *et al*. Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. *Protist* 2009;**160**:386–96.

Evans KM, Kühn SF, Hayes PK. High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudonitzschia pungens* (Bacillariophyceae) populations. *J Phycol* 2005;**41**:506–14.

Excoffier L, Lischer HEL. Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;**10**:564–7.

Fenchel T, Finlay BJ. The ubiquity of small species: Patterns of local and global diversity. *Bioscience* 2004;**54**:777.

Fernández LD, Fournier B, Rivera R *et al*. Water–energy balance, past ecological perturbations and evolutionary constraints shape the latitudinal diversity gradient of soil testate amoebae in south-western South America. *Global Ecol Biogeogr* 2016;**25**:1216–27.

Fernández LD, Lara E, Mitchell EAD. Checklist diversity and distribution of testate amoebae in Chile. *Eur J Protistol* 2015;**51**:409–24.

Finlay BJ. Global dispersal of free–living microbial eukaryote species. *Science* 2002;**296**:1061–3.

Foissner W. Dispersal and biogeography of protists: recent advances. *Jpn J Protozool* 2007;**40**:1–16.

Foissner W. Protist diversity and distribution: some basic considerations. *Biodivers Conserv* 2008;**17**:235–42.

Fournier B, Coffey EED, van der Knaap WO *et al*. A legacy of human-induced ecosystem changes: spatial processes drive the taxonomic and functional diversity of testate amoebae in *Sphagnum* peatlands of the Galápagos. *J Biogeogr* 2016;**43**:533–43.

Godhe A, Egardt J, Kleinhans D *et al*. Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. *Proc R Soc Lond* 2013;**280**:1–9.

Godhe A, Härnström K. Linking the planktonic and benthic habitat: genetic structure of the marine diatom *Skeletonema marinoi*. *Mol Ecol* 2010;**19**:4478–90.

González Díaz EF, Di Tommaso I. Paleogeoformas lacustres en los lagos Musters y Colhué Huapi, su relación genética con un paleolago Sarmiento previo, centro-sur del Chubut. *Rev Asoc Geol Argent* 2014;**71**:416–26.

Goslee SC, Urban DL. The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw* 2007;**22**:1–19.

Grant WAS, Bowen BW. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered* 1998;**89**:415–26.

Guillot G, Estoup A, Mortier F *et al*. A spatial statistical model for landscape genetics. *Genetics* 2005;**170**:1261–80.

Guillot G, Mortier F, Estoup A. Geneland: a computer package for landscape genetics. *Mol Ecol Notes* 2005;**5**:712–5.

Guillot G, Santos F, Estoup A. Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* 2008;**24**:1406–7.

Härnström K, Ellegaard M, Andersen TJ *et al*. Hundred years of genetic structure in a sediment revived diatom population. *P Natl Acad Sci USA* 2011;**108**:4252–7.

Izaguirre I, Saad JF, Schiaffino MR *et al*. Drivers of phytoplankton diversity in Patagonian and Antarctic lakes across a latitudinal gradient (2150 km): the importance of spatial and environmental factors. *Hydrobiologia* 2015;**764**:157–70.

Kim E, Wilcox L, Graham L *et al*. Genetically distinct populations of the dinoflagellate *Peridinium limbatum* in neighboring Northern Wisconsin lakes. *Microb Ecol* 2004;**48**:521–7.

Lara E, Fernández LD, Schiaffino RM *et al*. First freshwater member ever reported for the family Bathycoccaceae

(Chlorophyta; Archaeplastida) from Argentinean Patagonia revealed by environmental DNA survey. *Eur J Protistol* 2017;**60**:45–49.

Lara E, Heger TJ, Scheihing R *et al.* COI gene and ecological data suggest size-dependent high dispersal and low intra-specific diversity in free-living terrestrial protists (Euglyphida;Assulina). *J Biogeogr* 2011;**38**:640–50.

Lara E, Moreira D, Vereshchaka A *et al.* Pan–oceanic distribution of new highly diverse clades of deep–sea diplonemids. *Environ Microbiol* 2009;**11**:47–55.

Lee C-R, Mitchell-Olds T. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol Ecol* 2011;**20**:4631–42.

Lee-Yaw JA, Davidson A, Mcrae BH *et al.* Do landscape processes predict phylogeographic patterns in the wood frog? *Mol Ecol* 2009;**18**:1863–74.

Legendre P, Legendre L. *Numerical Ecology*, 2nd English edn. Amsterdam: Elsevier Science, 1998.

Librado P, Rozas J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;**25**:1451–2.

Lichstein JW. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol* 2007;**188**:117–31.

Logares R, Audic S, Bass D *et al.* Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* 2014;**24**:813–21.

Logares R, Boltovskoy A, Bensch S *et al.* Genetic diversity patterns in five protist species occurring in lakes. *Protist* 2009;**160**:301–17.

Lowe CD, Martin LE, Montagnes DJ *et al.* A legacy of contrasting spatial genetic structure on either side of the Atlantic-Mediterranean transition zone in a marine protist. *P Natl Acad Sci USA* 2012;**109**:20998–1003.

Mantel NA. The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967;**27**:209–20.

Mariazzi A, Conzonno VH, Ulibarrena J *et al.* Limnological investigation in Tierra del Fuego, Argentina. *Biol Acuat* 1987;**10**:1–74.

Marin B, Melkonian M. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* 2010;**161**:304–36.

Martiny JB, Bohannan BJ, Brown JH *et al.* Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006;**4**:102–12.

Muller MH, Leppälä J, Savolainen O. Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 2008;**100**:47–58.

Nagai S, Lian C, Shimada H *et al.* Microsatellite markers reveal population genetic structure of the toxic dinoflagellates *Alexandrium tamarense* (Dinophyceae) in Japanese coastal waters. *J Phycol* 2007;**43**:43–54.

Nosil P. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc R Soc Lond* 2004;**271**:1521–8.

Nosil P, Vines TH, Funk DJ. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 2005;**59**:705–19.

Ogden R, Thorpe RS. Molecular evidence for ecological speciation in tropical habitats. *P Natl Acad Sci USA* 2002;**99**:13612–5.

Palma RE, Boric-Barguetto D, Torres-Pérez F *et al.* Glaciation effects on the phylogeographic structure of *Oligoryzomys longicaudatus* (Rodentia: Sigmodontinae) in the Southern Andes. *PLoS One* 2012;**7**:e32206.

Paruelo JM, Beltrán A, Jobbágy E *et al.* The climate of Patagonia: general patterns and controls on biotic processes. *Ecol Aust* 1998;**8**:85–101.

Paruelo JM, Golluscio RA, Jobbágy EG *et al.* Situación ambiental en la estepa patagónica. In: Brown A, Martínez Ortiz U, Acerbi M *et al.* (eds). *La Situación Ambiental Argentina*. Buenos Aires, Argentina: Fundación vida silvestre, 2005, 303–20.

Quirós R, Drago E. The environmental state of Argentinean lakes: an overview. *Lakes Reserv Res Manage* 1999;**4**:55–64.

Rodríguez-Martínez R, Rocap G, Salazar G *et al.* Biogeography of the uncultured marine picoeukaryote MAST-4: temperature-driven distribution patterns. *ISME J* 2013;**7**:1531–43.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013. http://www.R-project.org/.

Rengefors K, Logares R, Laybourn-Parry J. Polar lakes may act as ecological islands to aquatic protists. *Mol Ecol* 2012;**21**:3200–9.

Rice WR. Analyzing tables of statistical test. *Evolution* 1989;**43**:223–5.

Rynearson TA, Newton JA, Armbrust EV. Spring bloom development, genetic variation, and population succession in the planktonic diatom *Ditylum brightwellii*. *Limnol Oceanogr* 2006;**51**:1249–61.

Schiaffino RM, Lara E, Fernández LD *et al.* Microbial eukaryote communities exhibit robust biogeographical patterns along a gradient of Patagonian and Antarctic lakes. *Environ Microbiol* 2016;**18**:5249–64.

Schiaffino RM, Unrein F, Gasol JM *et al.* Bacterial community structure in a latitudinal gradient of lakes: the roles of spatial versus environmental factors. *Freshwater Biol* 2011;**56**:1973–91.

Simmons MP, Sudek S, Monier A *et al.* Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the Eastern North Pacific Ocean. *Appl Environ Microb* 2016;**82**:1693–705.

Sjöqvist C, Godhe A, Jonsson PR *et al.* Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea-Baltic Sea salinity gradient. *Mol Ecol* 2015;**24**:2871–85.

Šlapeta J, Lopez-Garcia P, Moreira D. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Ecol* 2006;**23**:23–29.

Tell G, Izaguirre I, Allende L. Diversity and geographic distribution of Chlorococcales (Chlorophyceae) in contrasting lakes along a latitudinal transect in Argentinean Patagonia. *Biodivers Conserv* 2011;**20**:703–27.

Thorpe RS, Surget-Groba Y, Johansson H. The relative importance of ecology and geographic isolation for speciation in anoles. *Philos T Roy Soc B* 2008;**363**:3071–81.

Unrein F, Izaguirre I, Massana R *et al.* Nanoplankton assemblages in maritime Antarctic lakes: characterisation and molecular fingerprinting comparison. *Aquatic Microb Ecol* 2005;**40**:269–82.

Van den Wyngaert S, Möst M, Freimann R *et al.* Hidden diversity in the freshwater planktonic diatom *Asterionella formosa*. *Mol Ecol* 2015;**24**:2955–72.

Van Gremberghe I, Leliaert F, Mergeay J *et al.* Lack of phylogeographic structure in the freshwater cyanobacterium *Microcystis aeruginosa* suggests global dispersal. *PLoS One* 2011;**6**:e19561.

Vilela CR. Descripción Geológica de la Hoja 48c "Lago Musters", Provincia de Chubut. *Bol Inf Dir Nac Geol Minería* 1971;**113**:1–63.

Vucetich MC, Lopretto EC. Rhizopoda: Amebas testáceas. In: de Castellanos ZA, Coscarón S, Miquel S (eds). *Fauna de agua dulce de la República Argentina 3, Protozoa*. La Plata, Argentina: Facultad de Ciencias Naturales y Museo, 1995, 5–54.

Vyverman W, Verleyen E, Sabbe K *et al*. Historical processes constrain patterns in global diatom diversity. *Ecology* 2007;**88**:1924–31.

Wang IJ. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 2013;**67**:3403–11.

Wang IJ, Summers K. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol* 2010;**19**:447–58.

Weisse T. Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. *Biodivers Conserv* 2008;**17**:243–59.

Wilkinson DM. What is the upper size limit for cosmopolitan distribution in free–living microorganisms? *J Biogeogr* 2001;**28**:285–91.

Wilkinson DM, Koumoutsaris S, Mitchell EAD *et al*. Modelling the effect of size on the aerial dispersal of microorganisms. *J Biogeogr* 2012;**39**:89–97.

Woelfl S. Notas sobre protozoos ciliados dulceacuicolas de Chile. *Gayana* 2006;**70**:24–26.

Wright S. Isolation by distance. *Genetics* 1943;**28**:114–38.

Yang J, Smith HG, Sherrat TN *et al*. Is there a size limit for cosmopolitan distribution in free–living microorganisms? A biogeographical analysis of testate amoebae from polar areas. *Microb Ecol* 2010;**59**:635–45.