MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 5172–5184 doi: 10.1111/mec.12449

Holarctic phylogeography of the testate amoeba *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation

THIERRY J. HEGER,* EDWARD A. D. MITCHELL† and BRIAN S. LEANDER*

Departments of Botany and Zoology, University of British Columbia, #3529- 6270 University Blvd., Vancouver, British Columbia, Canada V6T 1Z4, †Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, Neuchâtel, CH-2000, Switzerland

Abstract

Although free-living protists play essential roles in aquatic and soil ecology, little is known about their diversity and phylogeography, especially in terrestrial ecosystems. We used mitochondrial cytochrome c oxidase subunit 1 (COI) gene sequences to investigate the genetic diversity and phylogeography of the testate amoeba morphospecies *Hyalosphenia papilio* in 42 *Sphagnum* (moss)-dominated peatlands in North America, Europe and Asia. Based on ≥1% sequence divergence threshold, our results from single-cell PCRs of 301 individuals revealed 12 different genetic lineages and both the general mixed Yule-coalescent (GMYC) model and the automatic barcode gap discovery (ABGD) methods largely support the hypothesis that these 12 *H. papilio* lineages correspond to evolutionary independent units (i.e. cryptic species). Our data also showed a high degree of genetic heterogeneity within different geographical regions. Furthermore, we used variation partitioning based on partial redundancy analyses (pRDA) to evaluate the contributions of climate and dispersal limitations on the distribution patterns of the different genetic lineages. The largest fraction of the variation in genetic lineage distribution was attributed to purely climatic factors (21%), followed by the joint effect of spatial and bioclimatic factors (13%), and a purely spatial effect (3%). Therefore, these data suggest that the distribution patterns of *H. papilio* genetic lineages in the Northern Hemisphere are more influenced by climatic conditions than by dispersal limitations.

Keywords: amoebozoa, biogeography, *Hyalosphenia papilio*, microbial eukaryotes, phylogeography, protists

Received 26 March 2013; revision received 26 June 2013; accepted 1 July 2013

Introduction

Free-living microbial eukaryotes or protists play fundamental roles in nutrient cycling and food-web processes in soils and aquatic environments (Adl & Gupta 2006; Wilkinson 2008); however, little is known about their overall diversity and phylogeography. Because of their large population sizes and small body sizes, dispersal limitation in free-living protists has been inferred to be essentially nonexistent, leading to opinions that these species tend to have cosmopolitan distribution patterns (Finlay 2002). This view, known as the 'ubiquity hypothesis', is supported by surveys of protist morphospecies and by some molecular phylogenetic data (Finlay & Clarke 1999; Slapeta *et al.* 2006; Pawlowski *et al.* 2007). However, several reports of morphospecies with restricted distributions challenge the validity of the 'ubiquity hypothesis' (Smith & Wilkinson 2007; Vanormelingen *et al.* 2008; Heger *et al.* 2011). Furthermore, environmental PCR surveys of eukaryotic microbial diversity have shown that the genetic variation

Correspondence: Thierry J. Heger, Fax: +1 604-822-6089; E-mail: thierry.heger@botany.ubc.ca

in some free-living protists corresponds with distinct spatial patterns. For instance, Lowe *et al.* (2010) demonstrated that some genetic lineages of the dinoflagellate *Oxyrrhis marina* occurred only in the Mediterranean sea; Bass *et al.* (2007) reported that some cercomonad species show restricted and/or patchy distributions; and Boo *et al.* (2010) reported that most phylotypes of the freshwater alga *Synura petersenii* have distinct geographical boundaries. More recently, Zufall *et al.* (2013) reported high levels of population subdivision, low rates of migration and significant isolation by distance in the freshwater aquatic ciliate *Tetrahymena thermophila*. There is therefore compelling evidence for biogeographical patterns in different groups of marine and freshwater free-living protists, but so far, molecular phylogenetic data from soil habitats such as peat soils remain very scarce (Epstein & López-García 2008; Fontaneto 2011). The processes that govern the diversity and distribution of protists have also received little attention, and it remains unclear whether biogeographical patterns in free-living protists result from historical processes (e.g. dispersal limitations) or contemporary environmental processes (e.g. local selection). Hanson *et al.* (2012) reviewed 54 studies that addressed this question and concluded that contemporary selection had a greater effect on microbial composition than historical processes. To date, no study has investigated the relative contribution of dispersal effects and environmental factors on the genetic lineages of a protist morphospecies or a protist species complex using a genetic approach such as DNA barcoding.

The main goals of this study were to examine the phylogeographical patterns and processes of *H. papilio* morphospecies, a putative species complex (Kosakyan *et al.* 2012), in Northern Hemisphere peatlands by addressing the following questions: (i) What is the extent of genetic variation within *H. papilio* morphospecies? (ii) How many distinct genetic lineages of *H. papilio* are there? (iii) Can these lineages be considered as evolutionary independent units (i.e. cryptic species)? (iv) To what extent do climatic and spatial factors explain the geographical patterns of genetic lineages of *H. papilio*? This morphospecies provides an excellent model for investigating the biogeography of free-living terrestrial protists because *H. papilio* is characterized by a very distinctive morphology: a large proteinaceous shell (120 μm) and the presence of intracellular *Chlorella*-like symbionts (Fig. 1), (Charret 1964; Ogden & Hedley 1980). These features facilitate the identification and isolation of this morphospecies for single-cell (SC) PCR and molecular phylogenetic analysis. *H. papilio* is also abundant in Northern Hemisphere *Sphagnum*-dominated peatlands, and its occurrence has been extensively documented in numerous ecological and palaeoecological studies (Meisterfeld 2002; Booth & Zygmunt 2005; Mitchell *et al.* 2008). By contrast, *H. papilio* appears to be lacking in Southern Hemisphere *Sphagnum*-dominated peatlands (Smith *et al.* 2008).

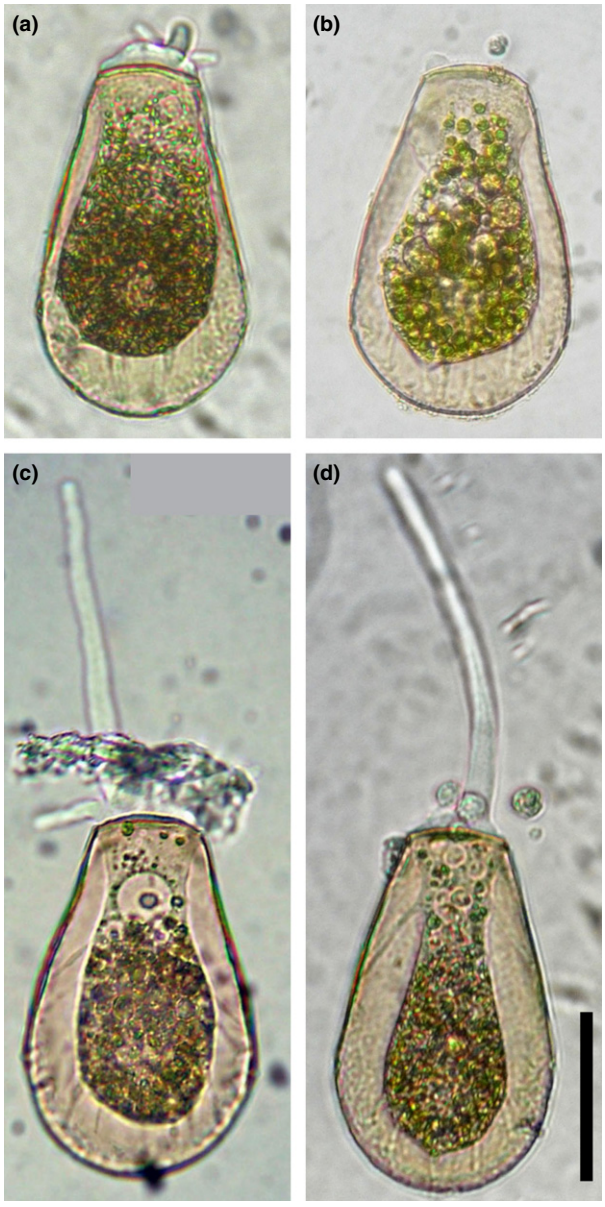

Fig. 1 Light micrographs (LM) illustrating four specimens of *Hyalosphenia papilio*. We isolated and extracted DNA from for single-cell PCR. Specimens shown in (a) and (c) were sampled in Alaska and belong to genetic lineage C. Specimens shown in (b) and (d) were sampled in Switzerland and belong to genetic lineage A. Scale bar = 60 μm.

Sphagnum-dominated peatlands are characterized by low pH and low concentrations of mineral nutrients, which act as strong environmental filters on communities. Although *Sphagnum*-dominated peatlands are not

highly biodiverse, they do harbour many characteristic species of plants, animals, fungi and microorganisms. As a result, these ecosystems often constitute isolated 'islands' in terrestrial landscapes. Many species, such as the testate amoeba *H. papilio*, are restricted to *Sphagnum*-dominated peatlands (Rydin & Jeglum 2006; Opelt *et al.* 2007a,b; Bragina *et al.* 2012). Thus, the island-like nature of *Sphagnum*-dominated peatlands offers a useful natural system for investigating the importance of dispersal limitation and environmental factors on the biogeographical patterns of terrestrial protists.

We used the mitochondrial cytochrome c oxidase subunit 1 (COI) gene to investigate the genetic diversity within *H. papilio* morphospecies on a global scale. COI was shown to be the marker of choice for assessing the genetic diversity and population structure in animals and several groups of protists, including ciliates, dinoflagellates, euglyphid testate amoebae, arcellinid testate amoebae and naked amoebae (Hebert *et al.* 2003; Gentekaki & Lynn 2009; Nassonova *et al.* 2010; Stern *et al.* 2010; Lara *et al.* 2011; Kosakyan *et al.* 2012; Zufall *et al.* 2013). The high number of mitochondria per cell facilitates PCR amplification of COI sequences from single-cell isolates. COI sequences are also characterized by a higher rate of sequence variation than nuclear small subunit (SSU) rDNA sequences, which facilitates high genetic resolution and allows phylogeographical investigations. We then assessed if climatic, spatial and micro-environmental factors explained the geographical patterns of *H. papilio* genetic lineages.

Materials and methods

Sampling and single-cell isolation

A total of 298 single cells of the morphospecies *Hyalosphenia papilio* were isolated from *Sphagnum* samples collected in 42 peatlands in North America, Europe and Asia (Table 1). *Hyalosphenia papilio* was extracted from the first five centimetres of the *Sphagnum* mosses by sieving and back-sieving using a 30- and 100-μm mesh. Individual cells were isolated using a narrow diameter pipette under an inverted microscope and washed three times with distilled water before being deposited into a 0.2-mL Eppendorf tube.

Single-cell PCR amplification and DNA sequencing

Mitochondrial cytochrome c oxidase subunit I gene (COI) fragments of 629 bp were amplified by seminested PCR using general COI primers LCO1490 and HCO2198 (Folmer *et al.* 1994) in the first reaction followed by a specific *H. papilio* primer (Arcox3R: 5′- ATA AAT GCT GAT ACA AAA TAG G3′) paired with the general LCO1490 primer in a second reaction. Single cells were added directly to a tube containing the general COI primers and an illustra PuReTaq Ready-to-Go PCR bead (GE Healthcare) in a final volume of 25 μL of water. The PCR cycling profile for the first PCR was as follows: an initial denaturation step for 4 min at 95 °C, followed by 39 cycles of 0.5 min of denaturation at 95 °C, 0.5 min of annealing at 42 °C and 2 min of elongation at 72 °C, with an additional final elongation step of 5 min at 72 °C. The cycling profile was the same for the second PCR in the seminested protocol, except that the annealing temperature was increased to 50 °C. Negative controls were used in all amplification steps. Each reaction was checked for length by running 3 μL of the PCR product on 2% agarose gels with 1× nucleic acid stain GelRed (Biotum Inc.). The PCR products were purified with exonuclease I/shrimp alkaline phosphatase (ExoSAP-IT, USB Corporation) and subsequently sequenced in both directions with a BigDye Terminator Kit. New COI sequences were manually edited in Bioedit 7.0.9 (Hall 1999) and identified with BLAST and molecular phylogenetic analysis before being deposited in GenBank (KC170411 - KC170708).

Multiple sequence alignment and molecular phylogenetic analyses

Molecular phylogenetic analyses were performed on a 301-taxon alignment consisting of 629 bp (298 new COI sequences as well as three previously published sequences from *H. papilio*) (Table 1). Eight previously published *H. papilio* sequences were not included in our analyses because they were either too short or not assignable to an unambiguous number of specimens because they were obtained from a multiple-cell DNA extraction. Inferred trees were rooted with COI sequences from two closely related species, *Nebela penardiana* (JN849062) and *Nebela galeata* (JN849060, JN849058 and JN849059) (Kosakyan *et al.* 2012). Trees were inferred with maximum-likelihood (ML) analyses using Treefinder (Jobb *et al.* 2004) and Bayesian inference methods using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). The Modeltest program (Posada & Crandall 1998) identified the TrN+I+ G as the most appropriate model of sequence evolution. ML analyses were run for 1000 replicates, and the most likely tree was chosen from those runs. Bayesian analyses consisted of three simultaneous Markov chains ran for 10 million generations from a random starting tree; trees were sampled every 1000 generations. The first 250 000 'burn-in' trees were discarded after checking that the chains had converged. The resultant trees were used to calculate the posterior probabilities (PP) for each node. Trees were viewed using FigTree (program distributed as part of BEAST: http://tree.bio.ed.ac.uk/software/figtree/).

Table 1 Brief description of the 42 *Sphagnum*-dominated peatland sampling sites

Sample site number	Latitude	Longitude	Country	Region	Sampling location	Altitude (m)	pH	Number of sequences per site	GenBank number
1	60.9852	−150.4267	USA	Alaska	Vogel Lake near Anchorage, Alaska	34	6.02	3	KC170411–KC170413
2	60.9945	−150.4214	USA	Alaska	Vogel Lake near Anchorage, Alaska	36	6.25	3	KC170414–KC170416
3	61.1602	−149.746	USA	Alaska	Near Anchorage, Alaska	120	4.69	14	KC170417–KC170430
4	68.6289	−149.5776	USA	Alaska	Toolik, Alaska	762	—	6	KC170431–KC170436
5	63.1247	−148.5001	USA	Alaska	Denali Highway, Anchorage, Alaska	883	—	5	KC170437–KC170441
6	65.1986	−148.0498	USA	Alaska	Wickersham wet taiga, Alaska	745	4.77	2	KC170442–KC170443
7	61.8323	−147.3481	USA	Alaska	Near Knob lake, Alaska	872	6.12	4	KC170444–KC170447
8	51.6586	−128.14	Canada	West British Columbia	Calvert Island, British Columbia	16	4.09	24	KC170448–KC170471
9	50.758	−126.4662	Canada	West British Columbia	Echo Bay, British Columbia	88	—	2	JN849011, JN849016
10	49.0127	−125.6578	Canada	West British Columbia	Pacific Rim, British Columbia	22	4.20	1	KC170472
11	48.8119	−125.1279	Canada	West British Columbia	Bamfield, British Columbia	33	4.83	4	KC170473–KC170476
12	48.8125	−125.1249	Canada	West British Columbia	Bamfield, British Columbia	28	4.30	2	KC170477–KC170478
13	48.8211	−125.1021	Canada	West British Columbia	Calamity Lake, Bamfield, British Columbia	15	5.43*	11	KC170479–KC170489
14	48.8358	−125.0626	Canada	West British Columbia	Bamfield, British Columbia	111	4.50	4	KC170490–KC170493
15	48.8372	−125.0617	Canada	West British Columbia	Bamfield, British Columbia	107	4.69	12	KC170494–KC170505
16	49.3977	−123.2079	Canada	West British Columbia	Cypress, British Columbia	922	4.11	4	KC170506–KC170509
17	49.4275	−123.2064	Canada	West British Columbia	Near Unnecessry Mountain, British Columbia	1340	—	1	JN849013
18	50.3674	−122.499	Canada	South central British Columbia	Near lower Joffre Lake, British Columbia	1206	5.28	1	KC170510
19	50.3838	−122.4524	Canada	South central British Columbia	West of Duffey Lake, British Columbia	1219	5.36*	23	KC170511–KC170533
20	52.1582	−119.272	Canada	South central British Columbia	Blue River, British Columbia	675	5.59*	17	KC170534–KC170550
21	52.5478	−119.1145	Canada	South central British Columbia	Allan Creek, British Columbia	824	5.38*	25	KC170551–KC170575
22	52.44	−117.46	Canada	Canadian Rockies	Rockies, Alberta	1594	4.92	7	KC170576–KC170582
23	44.4958	−63.9168	Canada	American east coast	Peggys Cove, Nova Scotia	8	—	4	KC170583–KC170586
24	42.8672	−6.8155	Spain	Castilla y León	Turbera Puerto Ancares	1545	5.16	14	KC170587–KC170600
25	50.5227	−3.9549	UK	South of Devon	Dartmoor	352	—	2	KC170601–KC170602
26	46.5566	6.1639	Switzerland	Jura mountains	Praz-Rodet	1041	—	2	KC170603–KC170604
27	46.5482	6.2326	Switzerland	Jura mountains	Amburnex	1308	5.42	2	KC170605–KC170606
28	46.8416	6.4658	Switzerland	Jura mountains	Vraconnaz	1091	4.50	8	KC170607–KC170614
29	47.1789	6.7897	France	Jura mountains	Russey	867	—	7	KC170615–KC170621
30	46.1513	6.8155	Switzerland	Alps	Bretolet	1825	5.95	9	KC170622–KC170630
31	46.3964	7.0997	Switzerland	Alps	Col des Mosses	1446	—	13	KC170631–KC170643
32	46.6114	7.9693	Switzerland	Alps	Höchrachen mire, Grindelwald	1886	—	3	KC170644–KC170646
33	58.6229	12.215	Sweden	South of Sweden	Province Dalsland	167	—	6	KC170647–KC170652

Table 1 *Continued*

Sample site number	Latitude	Longitude	Country	Region	Sampling location	Altitude (m)	pH	Number of sequences per site	GenBank number
34	53.8179	16.5941	Poland	North of Poland	Kusowo mire	146	5.24	6	KC170653–KC170658
35	53.1885	18.3092	Poland	North of Poland	Linje Mire	93	5.40	2	KC170659–KC170660
36	42.6	23.2833	Bulgaria	West central Bulgaria	Vitosha	1850	—	5	KC170661–KC170665
37	59.4313	25.8508	Estonia	North of Estonia	South of Lahemaa National Park	91	—	2	KC170666–KC170667
38	61.8182	56.851	Russia	Pechora	Yaksha South	122	—	1	KC170668
39	61.9115	57.7462	Russia	Pechora	Yaksha, Pechora reserve	168	4.20	7	KC170669–KC170675
40	62.0907	58.2942	Russia	Pechora	Pechora reserve	180	4.26*	10	KC170676–KC170685
41	62.0904	58.4395	Russia	Pechora	Pechora	181	4.69	6	KC170686–KC170691
42	52.0099	104.6644	Russia	South of Siberia	Near Lake Baïkal	463	—	17	KC170692–KC170708

*Denoted average pH value based on 2 or 3 values.

Lineage delineation strategies and haplotype diversity

We identified genetic lineages based on ≥1% sequence divergence. This threshold has been selected as in amoebae, a recent study revealed a very low level of COI intrastrain polymorphism (≤0.5%) and a limited to high level of interspecies divergence (i.e. minimum ca. 1%) (Nassonova *et al.* 2010). Sequence divergences were calculated using the Kimura 2-parameter (Kimura 1980) in R, version 3.01 (R Development Core Team 2010), using the package 'ape', version 3.0-1 (Table S1, Supporting information). To further evaluate the number of independent evolving units within *H. papilio* (i.e. putative cryptic species), we used the general mixed Yule-coalescent (GMYC) model (Pons *et al.* 2006; Fontaneto *et al.* 2007) and the automatic barcode gap discovery (ABGD) method (Puillandre *et al.* 2012). The GMYC method allows the identification of independent evolutionary units (IEUs) by discriminating population and speciation processes (Pons *et al.* 2006; Monaghan *et al.* 2009) while the ABGD method separates DNA sequences based on an automatic procedure of barcode gap discovery. For the GMYC model, we first removed identical sequences, and then, an ultrametric tree was generated in BEAST 1.6.2 (Drummond & Rambaut 2007) under the TrN+I+G model with an uncorrelated lognormal relaxed molecular clock model (Drummond *et al.* 2006) and a coalescence tree prior. Two independent Markov chain Monte Carlo (MCMC) were run for 10 million generations and sampled every 1000 steps. Of 2.5 million samples were discarded as a burn-in. The GMYC analysis was performed with the splits package, version 1.0-18 (Ezard *et al.* 2009), in R, version 3.01 (R Development Core Team 2010). Both the single threshold (Pons *et al.* 2006) and multiple threshold (Monaghan *et al.* 2009) methods were evaluated on the data set. Chi-squared test was used to test whether or not the performance of these two models differed significantly. Because no significant differences were revealed (chi square = 1.142, d.f. = 3, $P = 0.767$), only the results obtained from the single threshold model were reported. The ABGD method was implemented using the web server of the program with the Kimura K80 model and the default parameters (http://wwwabi.snv.jussieu.fr/public/abgd)/. For each *H. papilio* lineage, basic measurements of genetic diversity were calculated using the DnaSP software (Rozas *et al.* 2003), including the number of sequences (Nseq), the number of haplotypes (Nhap), the number of polymorphic sites (S), the haplotype diversity (h) and the nucleotide diversity (π) (Table 2).

Explanatory variables and variation partitioning

Two groups of explanatory variables were assessed for their effects on the distribution patterns of *H. papilio*

Table 2 Summary details of the number of sequences (*N*seq), haplotypes (*N*hap), polymorphic sites (S) as well as the diversity value of haplotype (*h*) and nucleotide (π) for *Hyalosphenia papilio*

	Lineages												
	A	B	C	D	E	F	G	H	I	J	K	L	All sequences
N seq	70	7	11	63	10	1	19	11	63	40	4	2	301
*N*hap	16	1	6	5	1	1	8	1	10	8	3	1	49
S	9	0	5	4	0	0	8	0	8	7	4	0	103
h	0.878	0	0.836	0.448	0	0	0.889	0	0.770	0.600	0.833	0	
π	0.0027	0	0.0031	0.0008	0	0	0.0040	0	0.0030	0.0026	0.0034	0	

lineages. The first one consisted of the altitude variable, water pH and 19 climatic variables extracted from the WORLDCLIM database (Table S2, Supporting information). WORLDCLIM, version 1.4; www.worldclim.org (Hijmans *et al.* 2005) consists of a set of climate layers at a spatial resolution of arc seconds (approximately 1×1 km) that were obtained by interpolation of climate station records from 1950 to 2000. WORLDCLIM is the highest resolution global climate data set available. Generic data were first converted into ESRI ASCII files using DIVA-GIS 5.2 (Hijmans 2001) before being imported and converted into raster in ArcMap using ArcGIS 10 Desktop (ESRI, Redlands, CA, USA). The 19 bioclimatic variables used in this study were coded as follows: annual mean temperature (BIO1), monthly average of diurnal temperature range (BIO2), isothermality (BIO3), temperature seasonality (standard deviation $\times$ 100) (BIO4), max temperature of warmest month (BIO5), min temperature of coldest month (BIO6), temperature annual range (BIO7), mean temperature of the wettest quarter (BIO8), mean temperature of driest quarter (BIO9), mean temperature of warmest quarter (BIO10), mean temperature of coldest quarter (BIO11), annual precipitation (BIO12), precipitation of wettest month (BIO13), precipitation of driest month (BIO14), precipitation seasonality (coefficient of variation) (BIO15), precipitation of wettest quarter (BIO16), precipitation of driest quarter (BIO17), precipitation of warmest quarter (BIO18) and precipitation of coldest quarter (BIO19). Small differences between the climatic conditions at the sampling sites and the Worldclim climatic variables cannot be ignored, but for this study, which aims to assess the genetic diversity within *H. papilio* morphospecies from different climatic regions, Worldclim provides reliable climatic estimates. Water pH, standardized to 20 °C, was measured on water extracted from 27 *Sphagnum* samples. Sample volume was insufficient for the remaining samples (Table 1). The second group of explanatory variables consisted of spatial variables, which were generated through the use of principal coordinates of neighbour matrices (PCNM) analysis. This analysis is based on the calculation of a principal coordinate analysis (PCoA) from a truncated matrix of Euclidean distances among sampling sites (Borcard & Legendre 2002; Borcard *et al.* 2011). In this study, the eigenvectors associated with positive eigenvalues were used as explanatory variables in ordination analyses to explain the effect of spatial distance between sampling sites on the distribution patterns of *H. papilio* lineages. PCNM variables better represent the spatial variation than geographical coordinates or polynomials (Borcard & Legendre 2002; Borcard *et al.* 2011). The *H. papilio* genetic data matrix corresponding to the abundance of genetic lineages at each site was Hellinger transformed (Legendre & Gallagher 2001). Hellinger transformation is a pretransformation of species data used for RDA that avoid considering the absence of the lineages as a resemblance between communities (Carlson *et al.* 2010). For multivariate analysis, we excluded four sampling sites, from which only one specimen was sequenced. Redundancy analysis (RDA) was applied using ANOVA permutations tests (999 permutations) with forward selection to identify environmental and spatial variables that significantly explained parts of the variation in the lineage composition data. For this forward selection procedure, we used the doublestop criterion method described by Blanchet *et al.* (2008). Variation partitioning was then carried out using only the selected explanatory variables to assess the significance of the different fractions. The R^2 values were adjusted to correct for differences in the number of samples and differences in the number of independent variables in both groups of explanatory data sets (Peres-Neto *et al.* 2006). To evaluate the linear dependencies between selected environmental variables, the variables' variance inflation factors were assessed (Borcard & Legendre 2002; Oksanen 2013). All variance inflation factors revealed weak colinearity between selected variables (VIF values < 10). All analyses were performed in R, version 3.01 (R Development Core Team 2010), using the package 'PCNM', version 2.1-2, to generate 'PCNM' eigenvectors, the version 0.0-8 of

the 'packfor' package for the forward selection of explanatory variables, and the packages 'vegan', version 2.0-2, for variation partitioning and variance inflation factor analyses.

Results

Single cells of the morphospecies *Hyalosphenia papilio* were isolated from *Sphagnum* samples collected in 42 peatlands of the Northern Hemisphere between 2008 and 2011. We performed seminested PCR using general COI primers in the first reaction followed by a specific *H. papilio* primer to amplify a 629-bp fragment of the mitochondrial cytochrome c oxidase subunit I gene (COI). We generated partial COI sequences from 298 single cells. In addition, three sequences were obtained from a previous study (JN849011, JN849013 and JN849016).

Genetic diversity

We analysed the genetic variation within 629 bp of the COI gene for a total of 301 sequences from individual cells of *H. papilio* isolated from 42 *Sphagnum*-dominated peatlands in North America (179 specimens), Europe (105 specimens) and Asia (17 specimens) (Table 1; Fig. 2). For each sampling site, the number of sequenced specimens per site ranged from 1 to 25 (Table 1). Examination of the 629-bp alignment revealed high levels of genetic diversity; 49 distinct haplotypes were identified, and a total of 103 sites were polymorphic (Table 2). The maximum sequence divergence between *H. papilio* sequences was 11.6%.

Phylogenetic lineages and evolutionary independent units

Twelve distinct lineages of *H. papilio* (A–L) were identified based on a ≥1% sequence divergence threshold (Fig. 2). Each of these lineages comprised from 1 to 70 sequences (average 25.1) and 1 to 16 haplotypes (average 5.1). Trees were inferred with maximum-likelihood (ML) analyses and Bayesian inference methods. Both methods of phylogenetic reconstruction resulted in similar topologies and provided high posterior probability support (i.e. ≥0.98) for all lineages and strong maximum-likelihood statistical support for lineages B, E, G, H, K and L (bootstrap values ≥95), moderate statistical support for lineages A, I and J (bootstrap values = 80–94), and relatively weak statistical support for lineages C and D (bootstrap values = 65–75) (Fig. 2).

The general mixed Yule-coalescent (GMYC) analysis based on the single threshold model revealed a total of 12 *H. papilio* GMYC independent clusters (eight clusters and four singletons) with a confidence interval of 11–16.

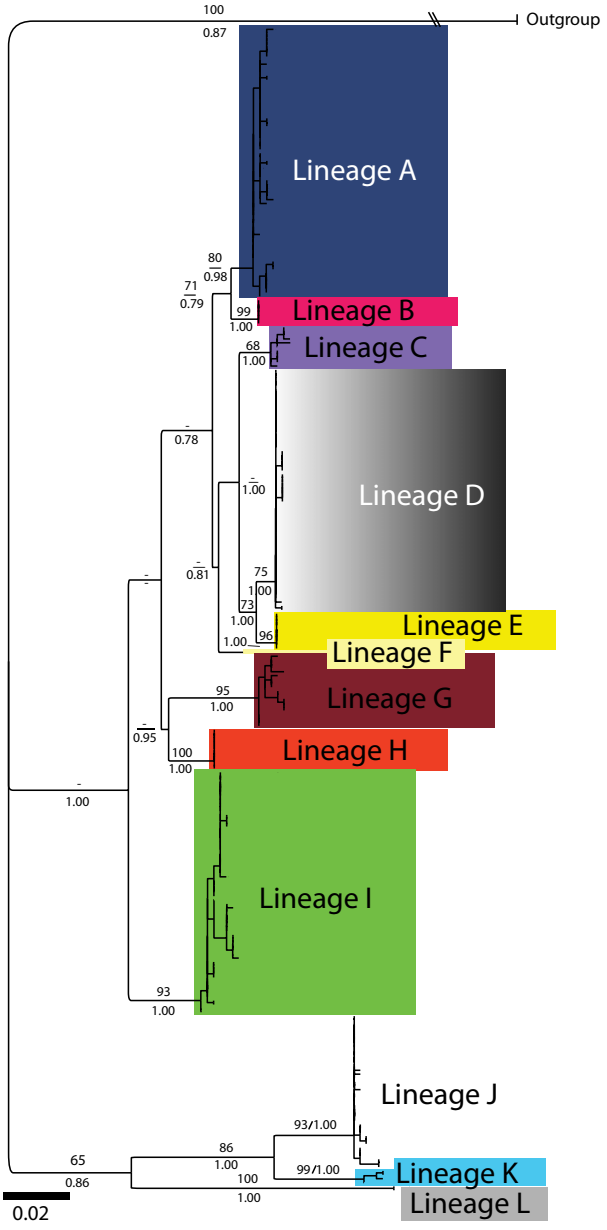

Fig. 2 Phylogenetic tree inferred from maximum-likelihood (ML) analysis of 629 nucleotide positions and 301 COI sequences from single-cell isolates of *Hyalosphenia papilio*. Numbers represent bootstrap values obtained by ML analysis (1000 replications) and posterior probabilities obtained by Bayesian analysis. Bootstrap values lesser than 65% are not shown. The tree was rooted with sequences from two arcellinid species: *Nebela penardiana* (JN849062) and *Nebela galeata* (JN849058, JN849059 and JN849060). Branch lengths are proportional to the number of substitutions per site (see scale bar). The number of sequences and haplotypes of each lineage are reported in Table 2.

This single threshold model provided a significantly better fit to the data than the null model of a single coalescent population (P = 0.035). These 12 evolutionary

independent units (EIUs) were the same as the ones identified based on a 1% delineation cut-off. The automatic barcode gap discovery (ABGD), based on standard settings, revealed seven a priori thresholds (i.e. barcoding gaps). Low a priori threshold values suggested 11 EIUs while higher threshold levels identified, respectively, 7 and 4 EIUs. Based on the fact that interspecies COI genetic divergence can be relatively low in amoebae (Nassonova *et al.* 2010), we considered the first scenario (i.e. 11 EIUs) more likely than the two other alternatives. These 11 EIUs defined by the ABGD method are also very congruent with the ones revealed by the ≥1% sequence divergence threshold, except the ABGD method grouped the clades D and E together.

Phylogeographical patterns

Our data showed a high degree of heterogeneity in genetic structure among different geographical regions (Fig. 3). Most lineages had restricted geographical distributions. For instance, lineages H and K were found only along the southwestern coast of British Columbia (Canada); lineage D was only found in South central British Columbia; lineage B was only found in the Canadian Rockies; lineage J was only found along the Canadian east coast near Halifax; and lineages E, G and O were only found in Alaska. Conversely, two lineages had a relatively widespread distribution: lineage F was found in Europe and North America (i.e. Pechora and Alaska), and lineage A was found in Asia and Europe (i.e. Palearctic) (Fig. 3). The number of genetic lineages per site varied from 1 to 4, but in most cases, only one lineage was found per site. Our results did not demonstrate any evidence of seasonal succession. We always found identical haplotypes when several samples were collected from the same location at different time points (e.g. sampling sites 16, 26 and 36).

Variation partitioning

Variation partitioning (Borcard *et al.*, 1992) was used to quantify the correlation of climatic and spatial variables with the distribution patterns of *H. papilio* lineages. Although the number of sequenced specimens varied from site to site, our data provided reliable estimations of *H. papilio* genetic lineage composition per site because few genetic lineages occurred at each site. Two groups of explanatory variables were used in this analysis: (1) altitude and 19 climatic variables and (2) spatial variables. To reduce the size of the two explanatory data sets, we ran forward selection using redundancy analyses (RDA). The four climatic variables selected were isothermality (BIO3), temperature seasonality (standard deviation × 100) (BIO4), mean temperature of warmest quarter (BIO10), mean temperature of coldest quarter (BIO11), annual precipitation (BIO12). For the second group of explanatory variables, only one spatial variable (i.e. PCNM1 eigenvector) was significantly correlated to compositional differences of *H. papilio* lineages. Variation partitioning was then carried out through a series

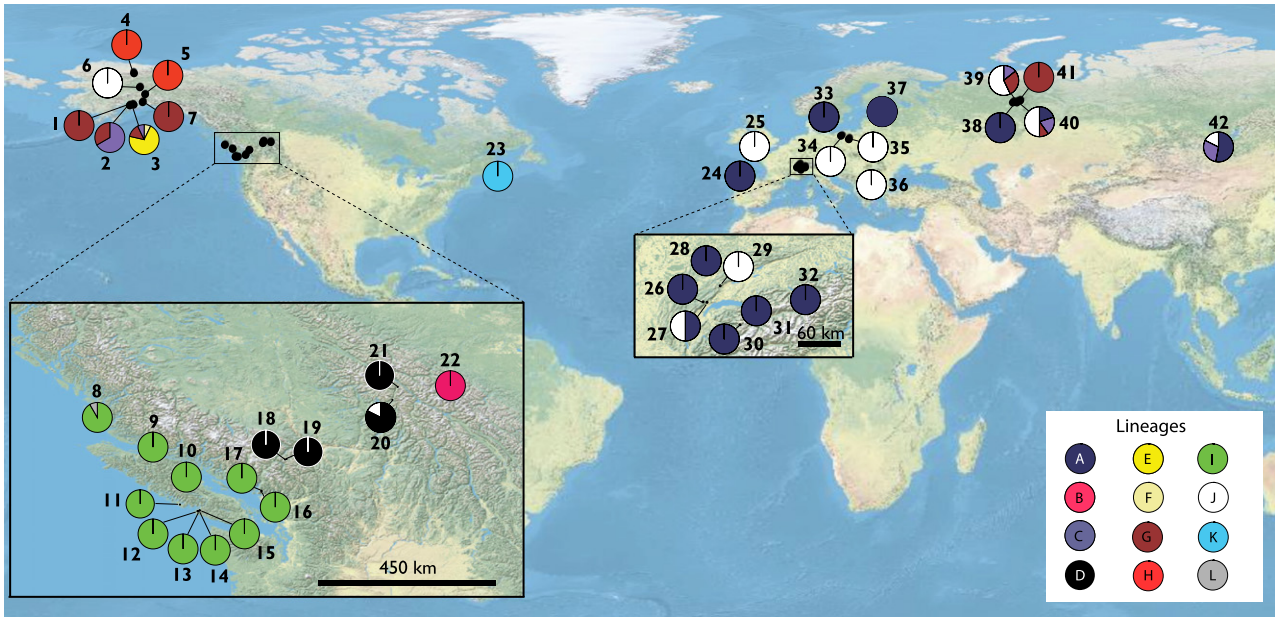

Fig. 3 Biogeographical distribution of the 301 single-cell isolates of the *Hyalosphenia papilio* morphospecies. Colours and letters indicate the 12 genetic lineages of *H. papilio* shown by phylogenetic analysis of the COI sequences. Pie charts refer to the proportion of sequences representing each genetic lineage at each sampling site (see Table 1). Colour assignments match those in Fig. 2.

of partial RDA (Borcard *et al.* 2011) using the five selected explanatory variables. Together, these five variables explained 37% of the total variation after correction for multiple comparisons (Peres-Neto *et al.* 2006). When this variation was decomposed, the largest fraction was attributed to purely climatic factors (21%), followed by the joint effect of spatial and climatic factors (13%), and a purely spatial effect (3%) (Fig. 3; in all cases $P < 0.05$, 999 permutations). In order to assess if our results were biased by the inclusion of sites from which only few sequences were obtained, we performed a second variation partitioning analysis in which sampling sites comprising less than four sequenced specimens were excluded (Fig. 4). The results of this analysis were almost identical to those of the full data set (i.e. the fraction of the variation attributed to purely climatic factors, the joint effect of spatial and the climatic factors and a purely spatial effect were, respectively, 21%, 12% and 4%). To further evaluate the impact of different sampling intensities among sites, we ran a complementary RDA analysis in which the total number of sequences per site was incorporated as a covariate. The forward selection indicated that this variable is not significantly correlated with compositional differences of *H. papilio* lineages. Finally, we also evaluated the correlation of pH with the compositional differences of *H. papilio* lineages by analysing a data set comprising only the sampling sites for which pH variables were available. Explanatory variables were the same as in the previous analyses, except that we included pH as an additional variable in the environmental group of explanatory variables. Similarly to the previous analyses, four climatic variables and one spatial variable were very significantly correlated with compositional differences of *H. papilio* lineages ($P < 0.01$ after 999 permutations), but no significant correlation was found between pH and *H. papilio* lineage distribution.

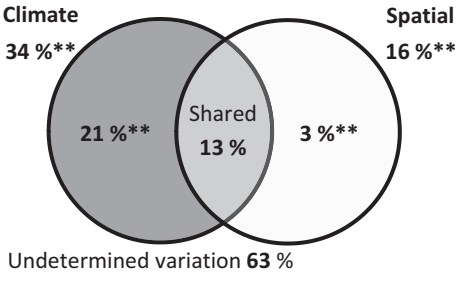

Fig. 4 Fractions of adjusted percent (R^2 adj) explained for the genetic composition of *Hyalosphenia papilio* by the set of predictor variables (climate = variation explained by climatic conditions and spatial = variation explained by spatial structure). ANOVA permutation tests were calculated on the variation explained by each set without the effect of the other (**$P < 0.01$).

Discussion

We used COI sequences to assess the genetic diversity and phylogeography of the testate amoeba morphospecies *Hyalosphenia papilio* from *Sphagnum*-dominated peatlands in the Northern Hemisphere. To our knowledge, this is the first global phylogeographical data set of COI sequences from over 300 single-cell isolates from a single protist morphospecies. This study also provides the first framework to quantify the relative influence of climatic and spatial factors on the genetic structure of a terrestrial protist morphospecies.

Genetic diversity and detailed phylogeography of H. papilio species complex

The combination of a high genetic resolution with sampling from three continents demonstrated a large level of genetic diversity with nonrandom geographical distributions. Most of the genetic lineages have been reported from only one continent. In Europe and along the Southwestern coast of British Columbia, where the sampling effort was the highest, only two different genetic lineages have been retrieved from each of these two regions, what supports the idea that not all lineages are cosmopolitan. Overall, 49 COI haplotypes were recovered from the 301 investigated specimens and 12 distinct lineages were found based on a 1% delineation cut-off. These results agree with our preliminary COI and SSU rRNA sequence data, which indicated a large degree of genetic diversity within *H. papilio* morphospecies (Kosakyan *et al.* 2012; T. Heger, unpublished results). Similar studies have shown moderate to high levels of genetic diversity within other major groups of protists; however, to our knowledge, such extensive COI molecular diversity seems higher than the COI diversity values reported from other studies. For instance, Lowe *et al.* (2010) found only four distinct haplotypes within the marine *Oxyrrhis marina* morphospecies isolated at broad geographical scale ($N = 58$); Gentekaki & Lynn (2012) reported 29 haplotypes distributed among six *Carchesium polypinum* clades from freshwater samples collected in the Grand River Basin (Canada), and a few other locations ($N = 100$); Zufall *et al.* (2013) identified 24 unique haplotypes in *Tetrahymena thermophila* ($N = 165$) isolated from ponds in the USA; and Lara *et al.* (2011) found, in European *Sphagnum*-dominated peatlands, only three slightly different haplotypes within the euglyphid testate amoeba *Assulina seminulum* ($N = 30$). However, the comparisons of these results should be made with caution, because the sampling intensity and the extent of the investigated geographical area vary between studies.

The maximum sequence divergence between lineages of *H. papilio* ranged from 1 to 11.6%, and even though no absolute threshold of COI sequence divergence has been established for delimiting species of either amoebae or any other protists (Boenigk *et al.* 2012), such high divergence values typically indicate the presence of different species (Kosakyan *et al.* 2012). In agreement with this hypothesis, the GMYC model recognized 12 evolutionary independent units (i.e. putative species) corresponding to the 12 lineages and the ABGD method provided almost the same conclusions. Therefore, we consider each of these 12 lineages as putative independent species. We chose not to redescribe these 12 different genetic lineages of *H. papilio* as new species because comparative morphology of the isolates using light microscopy (LM) did not provide any features that distinguish between members of the different lineages. For instance, variation in the number of shell pores did not correspond with the different lineages and can be interpreted as phenotypic plasticity, as previously suggested by Booth & Meyers (2010). However, further investigation of traits associated with the symbionts or ultrastructural systems might identify characteristic features for each of these 12 putative species.

Possible influence of environmental and spatial factors on the phylogeography of H. papilio

The variation partitioning analyses demonstrated that both environmental and spatial factors were significantly correlated with the geographical structure of genetic lineages in *H. papilio.* One-third of the total explained variation (13% of 37%) arose from the joint effect of environmental and spatial factors (i.e. climatic signal that is geographically structured). The fraction of variance explained by pure environmental factors was more than six times larger thanby pure spatial variables (21% vs. 3%). In other words, our data suggest that the distribution of *H. papilio* genetic lineages in the Northern Hemisphere is primarily driven by niche restrictions and secondarily by dispersal limitations. The data obtained along a strong climatic gradient in British Columbia and Alberta illustrate these conclusions. Sequences belonging to different lineages were found across three adjoining bioclimatic regions: the Pacific coastal region characterized by an oceanic climate, the interior region characterized by a continental climate and the Rockies region characterized by an alpine climate (Fig. 3). Interestingly, our results agree with other studies that found that, over a broad scale, local environmental factors have more impact on microbial communities than dispersal limitation (Hanson *et al.* 2012). Verleyen *et al.* (2009) performed one of the rare studies that quantified the relative importance of environmental and spatial factors on protist communities at a broad geographical scale. Their findings indicated that local environmental factors accounted for most of the explained variation in freshwater diatom communities, whereas dispersal-related factors were much less important.

In this study, we selected the four climatic factors that most influenced the assemblages of *H. papilio* genetic lineages: two corresponded to precipitation-related variables and two corresponded to temperature-related variables. These results were consistent with the findings of ecological studies suggesting that *H. papilio* responds primarily to moisture, and populations within geographically distinct peatlands have different moisture optima (which potentially correspond to the distinct phylotypes of *H. papilio*) (Booth 2001; Payne 2008). The importance of temperature on *H. papilio*-dominated communities has been documented by Jassey *et al.* (2011), but it is unknown if different *H. papilio* lineages (i.e. EIUs) have distinct temperature optima, as shown in other groups of protists. Souffreau *et al.* (2013), for instance, found that the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys* strains have lower optimal growth temperature and upper lethal temperature than most lineages from more temperate regions. Similarly, Boenigk *et al.* (2006) reported that phylotypes of '*Spumella*-like' flagellates isolated from different continents differ considerably with respect to their ecophysiology. Altogether, these data support the hypothesis that evolutionary independent units of *H. papilio* correlate with physiological tolerances. Interestingly, pH does not seem to be an important driver determining *H. papilio* genetic lineage distribution although this variable frequently emerges as significant in explaining the community patterns of testate amoeba morphospecies in peatlands (Mitchell *et al.* 2008).

The selected spatial PCNM variable 1 explained genetic variation that was not attributable to climatic factors. In other words, our data also provide support for the idea of dispersal limitation in the genetic lineage composition of *H. papilio* and thus for the so-called biogeography hypothesis or moderate endemicity distribution hypothesis, which assumes that at least some free-living protists have restricted geographical distributions (Foissner 2006). Although dispersal mechanisms in protists remain poorly understood, wind is thought to be one of the main agents for the dispersal of terrestrial testate amoeba (Wilkinson 2001). In addition, animals, including humans, represent potential vectors for the dispersal of protists (Foissner 2006; Perrigo *et al.* 2012). Dispersal limitations in terrestrial protists likely relate to the population size of a species, the ability to form resting cysts, the size of the test (if present) and

the spectrum of ecological tolerances (Finlay & Fenchel 2004; Mitchell & Meisterfeld 2005; Foissner 2008; Lara *et al.* 2011).

It is unlikely that population size explains the dispersal limitations of *H. papilio*, because this species is very abundant in Northern Hemisphere peatlands. Moreover, *H. papilio* has the ability to form resting cyst (Charret 1964) that probably facilitates its dispersal capability. By contrast, it is more likely that the comparatively large size of this species (about 120 μm, for both living and encysted individuals) limits its dispersal capability. Both empirical and modelling data have suggested that dispersal limitation increases with the size of the testate amoeba species, and the size of *H. papilio* is clearly in the range where passive dispersal over long distances is extremely unlikely (Wilkinson 2001; Yang *et al.* 2010; Wilkinson *et al.* 2012). Moreover, the narrow ecological tolerance of *H. papilio* might lower its rate of dispersal. If a species occurs only in relatively extreme and isolated habitats such as *Sphagnum*-dominated peatlands, it is more likely that dispersal limitation will impact its distribution. In this study, the spatial effect on the distribution of *H. papilio* genetic lineages has been interpreted as an indication of dispersal limitation. However, it is important to point out that we cannot completely rule out the possibility that the impact of dispersal limitation has been overestimated because a spatial signal can also emerge if unmeasured environment variation is spatially structured (Borcard *et al.* 2011). Although we included several relevant climatic and spatial factors in our analyses, 63% of the variation in *H. papilio* lineages assemblages remained unexplained. This variation might be explained by the substrate (*Sphagnum* species on which *H. papilio* lives) or by numerous others factors such as the ones known to directly or indirectly influence testate amoeba communities (e.g. UV-B, Ca^{2+}, NO_3^-) (Tolonen *et al.* 1994; Mitchell *et al.* 2000; Searles *et al.* 2001). In addition, a certain percentage of the unexplained variation might be related to the fact that the *H. papilio* diversity was not completely described, despite their important sampling effort. Obviously, extending the sampling and integrating additional environmental factors and conducting transplant experiments will be helpful to further advance our understanding of the influence of historical and contemporary environmental processes on the biogeography of *H. papilio*. Furthermore, additional investigations based on other protist species are needed in order to better understand soil protistan diversity and biogeography.

Acknowledgements

This work was funded by the Tula Foundation's Centre for Microbial Diversity and Evolution and the National Science and Engineering Research Council of Canada (NSERC 283091-09) to B.S.L., the Swiss NSF projects nr 205321–109709/1 and 205321–109709/2 to E.A.D.M. B.S. Leander is a Fellow of the Canadian Institute for Advanced Research. We would like to thank Enrique Lara for constructive discussions and help with single-cell isolation and sequencing, Alessia Guggisberg for several discussions about the analyses of the data, François Gillet for his comments on multivariate analyses, Julia Gustavsen for her comments on a preliminary version of the manuscript, Omar Fiz-Palacios for his help with the GMYC analysis and constructive discussion about the diversification of arcellinids, and all people who supplied us with samples and assisted with field sampling, notably L. Wiget Mitchell and kids, T. Asada, F. Burki, A, Buttler, M. Carlson, T. Cedhagen, J.C. and D. Heger, P. Keeling, L. and M. Lamentowicz, Y. Mazei, N. Okamoto, X. Pontevedra Pombal and colleagues, T. and J. Schwander, M. Todorov, B. Warner, D. Wilkinson, B. Yermen, and several members of the Leander laboratory. We also thank Diego Fontaneto and two anonymous referees for insightful comments. Finally, we thank the staff of the Bamfield Marine Sciences Centre and the Haikai Beach Institute (B.C., Canada).

References

Adl SM, Gupta VVSR (2006) Protists in soil ecology and forest nutrient cycling. *Canadian Journal of Forest Research*, **36**, 1805–1817.

Bass D, Richards TA, Matthai L, Marsh V, Cavalier-Smith T (2007) DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evolutionary Biology* **7**, 162.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology*, **89**, 2623–2632.

Boenigk J, Pfandl K, Garstecki T *et al.* (2006) Evidence for geographic isolation and signs of endemism within a protistan morphospecies. *Applied and Environmental Microbiology*, **72**, 5159–5164.

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.

Boo SM, Kim HS, Shin W *et al.* (2010) Complex phylogeographic patterns in the freshwater alga *Synura* provide new insights into ubiquity vs. endemism in microbial eukaryotes. *Molecular Ecology*, **19**, 4328–4338.

Booth RK (2001) Ecology of testate amoebae (Protozoa) in two lake superior coastal wetlands: implications for paleoecology and environmental monitoring. *Wetlands*, **21**, 564–576.

Booth RK, Meyers B (2010) Environmental controls on pore number in *Hyalosphenia papilio*: implications for paleoenvironmental reconstruction. *Acta Protozoologica*, **49**, 29–35.

Booth RK, Zygmunt JR (2005) Biogeography and comparative ecology of testate amoebae inhabiting *Sphagnum*-dominated peatlands in the Great Lakes and Rocky Mountain regions of North America. *Diversity and Distributions*, **11**, 577–590.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology With R*, 1st edn. Springer, New York.

Bragina A, Berg C, Cardinale M *et al.* (2012) *Sphagnum* mosses harbour highly specific bacterial diversity during their whole lifecycle. *ISME Journal*, **6**, 802–813.

Carlson ML, Flagstad LA, Gillet F, Mitchell EAD (2010) Community development along a proglacial chronosequence: are above-ground and below-ground community structure controlled more by biotic than abiotic factors? *Journal of Ecology*, **98**, 1084–1095.

Charret R (1964) Contribution à l'étude cytologique et biologique de *Hyalosphenia papilio* (Leidy), Rhizopode Testacé. *Bulletin Biologique de la France et de la Belgique*, **XCVIII**, 369–390.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Plos Biology*, **4**, 699–710.

Epstein S, López-García P (2008) "Missing" protists: a molecular prospective. *Biodiversity and Conservation*, **17**, 261–276.

Ezard T, Fujisawa T, Barraclough TG (2009) SPecies LImits by Threshold Statistics. http://r-forge.r-project.org/projects/splits/.

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061–1063.

Finlay BJ, Clarke KJ (1999) Ubiquitous dispersal of microbial species. *Nature*, **400**, 828.

Finlay BJ, Fenchel T (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist*, **155**, 237–244.

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozoologica*, **45**, 111–136.

Foissner W (2008) Protist diversity and distribution: some basic considerations. *Biodiversity and Conservation*, **17**, 235–242.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fontaneto D (2011) *Biogeography of Microscopic Organisms*. The Systematics Association edn. Cambridge University Press, Cambridge.

Fontaneto D, Herniou EA, Boschetti C *et al.* (2007) Independently evolving species in asexual bdelloid rotifers. *Plos Biology*, **5**, 914–921.

Gentekaki E, Lynn DH (2009) High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the Peritrichous Ciliate *Carchesium polypinum* in the Grand River Basin (North America). *Applied and Environmental Microbiology*, **75**, 3187–3195.

Gentekaki E, Lynn DH (2012) Spatial genetic variation, phylogeography and barcoding of the peritrichous ciliate *Carchesium polypinum*. *European Journal of Protistology*, **48**, 305–313.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, **10**, 497–506.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Heger TJ, Booth RK, Sullivan ME *et al.* (2011) Rediscovery of *Nebela ansata* (Protista: Arcellinida) in the eastern North-America– Biogeographical implications. *Journal of Biogeography*, **38**, 1897–1906.

Hijmans RJ (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* **127**, 15–19.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Jassey VEJ, Gilbert D, Binet P, Toussaint ML, Chiapusio G (2011) Effect of a temperature gradient on *Sphagnum fallax* and its associated living microbial communities: a study under controlled conditions. *Canadian Journal of Microbiology*, **57**, 226–235.

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* **4**, 18.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kosakyan A, Heger TJ, Leander BS *et al.* (2012) COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the Hyalospheniidae Schultze. *Protist*, **163**, 415–434.

Lara E, Heger TJ, Scheihing R, Mitchell EAD (2011) COI gene and ecological data suggest size-dependent high dispersal and low intra-specific diversity in free-living terrestrial protists (Euglyphida: Assulina). *Journal of Biogeography*, **38**, 640–650.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Lowe CD, Montagnes DJS, Martin LE, Watts PC (2010) Patterns of genetic diversity in the marine heterotrophic flagellate *Oxyrrhis marina* (Alveolata: Dinophyceae). *Protist*, **161**, 212–221.

Meisterfeld R (2002) Order Arcellinida Kent, 1880. In: *The Illustrated Guide to the Protozoa* (eds Lee JJ, Leedale GF & Bradbury P), pp. 827–860. Society of Protozoologists, Lawrence, Kansas.

Mitchell EAD, Meisterfeld R (2005) Taxonomic confusion blurs the debate on cosmopolitanism versus local endemism of free-living protists. *Protist*, **156**, 263–267.

Mitchell EAD, Buttler A, Grosvernier P *et al.* (2000) Relationships among testate amoebae (Protozoa), vegetation and water chemistry in five *Sphagnum*-dominated peatlands in Europe. *New Phytologist*, **145**, 95–106.

Mitchell EAD, Charman DJ, Warner BG (2008) Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. *Biodiversity and Conservation*, **17**, 2115–2137.

Monaghan MT, Wild R, Elliot M *et al.* (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

Nassonova E, Smirnov A, Fahrni J, Pawlowski J (2010) Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist*, **161**, 102–115.

Ogden CG, Hedley RH (1980) *An Atlas of Freshwater Testate Amoebae*. British Museum (Natural History) and Oxford University Press, London and Oxford.

Oksanen J (2013) Multivariate analysis of ecological communities in R: vegan tutorial.

Opelt K, Berg C, Schonmann S, Eberl L, Berg G (2007a) High specificity but contrasting biodiversity of *Sphagnum*-associated bacterial and plant communities in bog ecosystems independent of the geographical region. *ISME Journal*, **1**, 502–516.

Opelt K, Chobot V, Hadacek F *et al.* (2007b) Investigations of the structure and function of bacterial communities associated with *Sphagnum* mosses. *Environmental Microbiology*, **9**, 2795–2809.

Pawlowski J, Fahrni J, Lecroq B *et al.* (2007) Bipolar gene flow in deep-sea benthic foraminifera. *Molecular Ecology*, **16**, 4089–4096.

Payne RJ (2008) Testate amoebae as palaeohydrological proxies in Sürmene Agacbasi Yaylasi Peatland (Northeast Turkey). *Wetlands*, **28**, 311–323.

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.

Perrigo AL, Romeralo M, Baldauf SL (2012) What's on your boots: an investigation into the role we play in protist dispersal. *Journal of Biogeography*, **39**, 998–1003.

Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, **21**, 1864–1877.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Version 2.8.0*. R Development Core Team, Vienna, Austria. http://www.R-project.org.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Rydin H, Jeglum JK (2006) *Biology of Peatlands*. Oxford University Press, Oxford.

Searles PS, Kropp BR, Flint SD, Caldwell MM (2001) Influence of solar UV-B radiation on peatland microbial communities of southern Argentinia. *New Phytologist*, **152**, 213–221.

Slapeta J, Lopez-Garcia P, Moreira D (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution*, **23**, 23–29.

Smith HG, Wilkinson DM (2007) Not all free-living microorganisms have cosmopolitan distributions - the case of *Nebela* (*Apodera*) *vas* Certes (Protozoa: Amoebozoa: Arcellinida). *Journal of Biogeography*, **34**, 1822–1831.

Smith HG, Bobrov A, Lara E (2008) Diversity and biogeography of testate amoebae. *Biodiversity and Conservation*, **17**, 329–343.

Souffreau C, Vanormelingen P, Van de Vijver B *et al.* (2013) Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist*, **164**, 101–115.

Stern RF, Horak A, Andrew RL *et al.* (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. *PLoS ONE* **5**, 11.

Tolonen K, Warner BG, Vasander H (1994) Ecology of testaceans (Protozoa: Rhizopoda) in Mires in Southern Finland: II. Multivariate-analysis. *Archiv für Protistenkunde*, **144**, 97–112.

Vanormelingen P, Verleyen E, Vyverman W (2008) The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, **17**, 393–405.

Verleyen E, Vyverman W, Sterken M *et al.* (2009) The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*, **118**, 1239–1249.

Wilkinson DM (2001) What is the upper size limit for cosmopolitan distribution in free-living microorganisms? *Journal of Biogeography*, **28**, 285–291.

Wilkinson DM (2008) Testate amoebae and nutrient cycling: peering into the black box of soil ecology. *Trends in Ecology & Evolution*, **23**, 596–598.

Wilkinson DM, Koumoutsaris S, Mitchell EAD, Bey I (2012) Modelling the effect of size on the aerial dispersal of microorganisms. *Journal of Biogeography*, **39**, 89–97.

Yang Y, Smith HG, Sherratt TN, Wilkinson DM (2010) Is there a size limit for cosmopolitan distribution in free-living microorganisms? A biogeographical analysis of testate amoebae from polar areas. *Microbial Ecology*, **59**, 635–645.

Zufall RA, Dimond KL, Doerder FP (2013) Restricted distribution and limited gene flow in the model ciliate Tetrahymena thermophila. *Molecular Ecology*, **22**, 1081–1091.

T.J.H. and E.A.D.M. designed the study. T.J.H. and E.A.D.M. sampled material. T.J.H. performed sequencing and analysed the data. All authors discussed the results and wrote the manuscript.

Data accessibility

DNA sequences: For GenBank accession numbers and corresponding locations, see Table 1.

Phylogenetic data: TreeBASE Submission ID S14464.

Climatic data: Table S2.

Supporting information

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

Table S1 *Hyalosphenia papilio* sequence divergence matrix.

Table S2 Climatic data of the 42 sampling sites.