UNIVERSITY *of* York

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Dispersal limitations and historical factors determine the biogeography of specialized terrestrial protists

David Singer[1,2], Edward A.D. Mitchell[1,3], Richard J. Payne[4], Quentin Blandenier[1,5], Clément Duckert[1], Leonardo D. Fernández[1,6], Bertrand Fournier[7] Cristián E. Hernández[8], Gustaf Granath[9], Håkan Rydin[9], Luca Bragazza[10,11,12], Natalia G. Koronatova[13], Irina Goia[14], Lorna I. Harris[15], Katarzyna Kajukało[16], Anush Kosakyan[17] Mariusz Lamentowicz[16], Natalia P. Kosykh[13], Kai Vellak[18], Enrique Lara[1,5]

Corresponding author: David Singer, Department of Zoology, Institute of Biosciences, Rua do Matão, Tv. 14, n101, São Paulo, SP - 05508-090, Brazil, tel: 0041792182426, email: david.singer.bio@outlook.com

Affiliations:

[1]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

[2]Department of Zoology, Institute of Biosciences, University of São Paulo, 05508-090, Brazil

[3]Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

[4]Environment, University of York, Heslington, York, YO10 5DD, United Kingdom

[5]Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain

[6]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile

[7]Community and Quantitative Ecology Laboratory, Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal, QC H4B 1R6, Canada

[8]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile

[9]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.

[10]WSL Swiss Federal Institute for Forest, Snow and Landscape Research, Site Lausanne, Station 2, CH-1015 Lausanne, Switzerland

[11]Laboratory of Ecological Systems (ECOS), Ecole Polytechnique Féderale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Station 2, CH-1015 Lausanne, Switzerland

[12]Department of Life Science and Biotechnologies, University of Ferrara, Corso Ercole I d'Este 32, I-44121 Ferrara, Italy

[13]Laboratory of Biogeocenology, Institute of Soil Science and Agrochemistry, Siberian Branch of Russian Academy of Sciences, Ak. Lavrent'ev ave., 8/2, Novosibirsk, 630090 Russia

[14]Babeș-Bolyai University, Faculty of Biology and Geology, Department of Taxonomy and Ecology, 42 Republicii Street, RO-400015, Cluj-Napoca, Romania

[15]School of Geography and Earth Sciences, McMaster University, General Sciences Building, Hamilton, Ontario, L8S 4K1, Canada

[16]Laboratory of Wetland Ecology and Monitoring, Faculty of Geographical and Geological Sciences and Department of Biogeography and Paleoecology, Adam Mickiewicz Universi-ty, Poznań, Poland

[17] Institute of Parasitology, Biology Center, Czech Academy of Sciences, Branisovska 1160/31, 37005 Czeske Budejóvice, Czech Republic

[18]Institute of Ecology and Earth Sciences, Natural History Museum, University of Tartu, Lai St 40, Tartu 51005, Estonia

Abstract

44   Recent studies show that soil eukaryotic diversity is immense and dominated by

45   microorganisms. However, it is unclear to what extent the processes that shape the

46   distribution of diversity in plants and animals also apply to microorganisms. Major

47   diversification events in multicellular organisms have often been attributed to long-term

48   climatic and geological processes, but the impact of such processes on protist diversity has

49   received much less attention as their distribution has often been believed to be largely

50   cosmopolitan. Here, we quantified phylogeographic patterns in *Hyalosphenia papilio,* a large

51   testate amoeba restricted to Holarctic *Sphagnum*-dominated peatlands, to test if the current

52   distribution of its genetic diversity can be explained by historical factors or by the current

53   distribution of suitable habitat. Phylogenetic diversity was higher in Western North America,

54   corresponding to the inferred geographical origin of the *H. papilio* complex, and was lower in

55   Eurasia despite extensive suitable habitat. These results suggest that patterns of phylogenetic

56   diversity and distribution can be explained by the history of Holarctic *Sphagnum* peatland

57   range expansions and contractions in response to Quaternary glaciations that promoted

58   cladogenetic range evolution, rather than the contemporary distribution of suitable habitats.

59   Species distributions were positively correlated with climatic niche breadth, suggesting that

60   climatic tolerance is key to dispersal ability in *H. papilio.* This implies that, at least for large

61   and specialized terrestrial microorganisms, propagule dispersal is slow enough that historical

62   processes may contribute to their diversification and phylogeographic patterns and may partly

63   explain their very high overall diversity.

64   **Keywords**

65   *Hyalosphenia papilio*, phylogeography, *Sphagnum* peatland, protists, Distribution, Holarctic

Introduction

The question of whether the same rules structure the diversity of all eukaryotes, micro and – macroscopic alike, has been the subject of a heated debate since the early 2000's. The classical paradigm is that "everything is everywhere, but, the environment selects" (Baas-Becking, 1934). Defenders of this paradigm have argued that geographic barriers are ineffective in preventing dispersal of microbes (Fenchel, 2005; Finlay, 1998). Other researchers, while accepting that some microbes do indeed occur worldwide, have argued that others are clearly restricted to certain regions: the 'moderate endemicity' model (Foissner, 1999). This argument is based particularly on a limited number of so-called "biogeographic flagship species", with conspicuous morphology.

The application of barcoding to protists has brought new nuance to the debate (Pawlowski et al., 2012). Single-cell DNA barcoding studies (Pawlowski et al., 2012) of individual "morphospecies" are now revealing the existence of numerous "cryptic" biological species (Singer et al., 2018). Barcoding studies have now demonstrated geographically limited distributions in soil (Ryšánek, Hrčková, & Škaloud, 2015) freshwater (Škaloud et al., 2019) and marine organisms (Santoferrara, Rubin, & Mcmanus, 2018), although cases of cosmopolitan distribution have also been reported (Geisen, Fiore-Donno, Walochnik, & Bonkowski, 2014; Šlapeta, López-García, & Moreira, 2005). The development of microbial phylogeography, combining biogeography and molecular phylogeny, in turn allows the evaluation of possible drivers of diversity patterns, and comparison to those known to drive plant and animal diversity (Martiny et al., 2006).

Among terrestrial microorganisms, testate amoebae are particularly useful models for phylogeographical studies. Testate amoebae are conspicuous and relatively easy to identify

89 and are also large enough to be isolated individually for DNA barcoding. Many species have

90 narrow ecological tolerances and thus can only colonize specific, often geographically

91 discontinuous habitats (Singer et al., 2018). Furthermore, some species have well-documented

92 distribution and ecology. A good example is *Hyalosphenia papilio*, a widely recorded and

93 morphologically distinctive testate amoeba taxon (Fig. 1). Of particular interest for

94 phylogeographic studies is that, based on single cell barcoding and the variable molecular

95 marker Cytochrome c Oxidase subunit I (COI), *H. papilio* is known to represent a species

96 complex of at least twelve lineages (Heger, Mitchell, & Leander, 2013).

97 *Hyalosphenia papilio* is found exclusively in Holarctic *Sphagnum*-dominated peatlands

98 (Amesbury et al., 2018, 2016) and it is known to be absent from similar southern hemisphere

99 sites despite extensive study (Fernández, Lara, & Mitchell, 2015; Smith, Bobrov, & Lara,

100 2007). *Sphagnum*-dominated peatlands are comparatively young ecosystems, dating back to

101 the expansion of boreal and subarctic environments near the Pliocene (Shaw et al., 2010).

102 *Sphagnum* is an ecosystem engineer that modifies habitats by increasing soil wetness and

103 decreasing pH and available nutrient content, producing decay-resistant litter rich in phenols

104 and sphagnan (van Breemen, 1995) and hosting very distinctive prokaryotic, algal and fungal

105 communities (Kostka et al., 2016; Mutinová, Neustupa, Bevilacqua, & Terlizzi, 2016). Thus,

106 *Sphagnum* represents a highly selective habitat for macro and microorganisms. This explains

107 why *Sphagnum*-dominated ecosystems are species-poor and these same factors are likely to

108 also drive evolutionary adaptations in testate amoebae (Kosakyan et al., 2016; Singer et al.,

109 2018). Hence, it is likely that this taxon does not pre-date the radiation of peat-forming

110 *Sphagnum* species between 17 and 7 Mya (Shaw et al., 2010). Large extents of *Sphagnum*

111 comparable to modern *Sphagnum*-dominated peatlands probably appeared during the late

112 Miocene/early Pliocene, concomitantly with global cooling, i.e. between 7 and 5.5 Mya

4

(Herbert et al., 2016). While *Sphagnum* occurs at low as well as high latitudes it is only a

dominant component of peatlands in higher latitudes (e.g. Tierra-del-Fuego and the boreal

zone of the Holarctic). The taxonomic richness of the genus is low in the Southern

Hemisphere high latitudes and high in the Northern Hemisphere high latitudes which

correspond to its inferred origin (Shaw, Carter, Aguero, da Costa, & Crowl, 2019).

Holarctic *Sphagnum*-dominated peatlands have experienced considerable changes in their

extent due to the repeated advances and retreats of ice sheets during the Quaternary. Many of

the largest areas covered by peatlands today were under ice during the last glacial maximum

(e.g. Fennoscandia, boreal Canada), while peatlands may have persisted in others (e.g. Pacific

coast of Canada)(Treat et al., 2019). These successive glacial expansions and contractions are

known to have shaped genetic diversity in multicellular taxa (Schönswetter, Stehlik,

Holderegger, & Tribsch, 2005), whose dispersal is assumed to be slow in contrast with

eukaryotic microorganisms (Bahram et al., 2016). If, like plants and animals, protist dispersal

is relatively slow, the genetic structure of their populations will bear traces of such range

expansions and contractions. The origin of taxa can potentially be inferred and the timing of

phylogenetic events estimated based on molecular clocks (Arbogast et al., 2006; Arbogast,

Edwards, Wakeley, Beerli, & Slowinski, 2002). On the contrary, fast dispersal in protists

would blur any such signature and taxonomic or phylogenetic diversity would tend to be

distributed randomly and peak in areas with largest extent of favourable habitats (Forest,

Colville, & Cowling, 2018).

It follows that the phylogeographical pattern of a given taxon, here *H. papilio,* can be used to

test two alternative hypotheses: 1) Dispersal is low and/or slow enough so that traces of

glacial cycles are reflected in its extant diversity. The highest diversity, and the likely

geographical origin of *H. papilio* would be expected to occur in refugia corresponding to the

margins of ice sheets during Last Glacial Maximum where *Sphagnum* peatlands could

survive. 2) Dispersal is high and/or fast, and diversity would be expected to be maximal

where the largest expanses of *Sphagnum* peatlands are found today (e.g. Western Siberia).

Empirical evidence demonstrates a relationship between testate amoeba shell size and

geographic range (Wilkinson, 2001). Population genetics analyses (Lara, Heger, Scheihing, &

Mitchell, 2011) and modelling (Wilkinson, Koumoutsaris, Mitchell, & Bey, 2011) show a

decline in dispersal potential for testate amoebae and theoretical organisms of smaller sizes

(ca. 60 µm) than that reported for *H. papilio* (size range 90-175µm). The ability of entering a

dormant stage (cysts) which can withstand desiccation and other stresses is considered to be a

key dispersal trait in protists (Geisen et al., 2018); however, such structures have never been

reported in *H. papilio*. We therefore predict that the first hypothesis is more likely to be

supported.

Material and Methods

Dataset preparation

We retrieved all 360 existing COI gene sequences of *H. papilio* from GenBank, together with

information on the origin of the cells from four studies (Gomaa et al., 2014; Heger et al.,

2013; Kosakyan et al., 2012; Oliverio, Lahr, Grant, & Katz, 2015). In addition, we isolated 57

single cells of *H. papilio* from *Sphagnum* samples collected at 13 new sites, targeting under-

sampled regions to compile a global dataset (Supplementary Table 1). Briefly, the cells were

washed three times in autoclaved distilled water before DNA extraction, which was

performed using the guanidine thiocyanate-base protocol (Duckert et al., 2018).

Amplifications of COI gene fragments were performed in two steps: a first PCR was

160   undertaken with the general COI primers LCO1490 and HCO2198 (Folmer, Black, Hoeh,

161   Lutz, & Vrijenhoek, 1994), which was followed by a nested PCR using, *H. papilio*-specific

162   primers HPcoiF and HPcoiR (Gomaa et al., 2014). The first DNA amplification profiles

163   consisted of an initial denaturation step for 3 min at 95°C, followed by 39 cycles of 15 sec of

164   denaturation at 95°C, 15 sec of annealing at 43°C and 1 min of elongation at 72°C with an

165   additional final elongation step at 72°C for 10 min. The procedure for the second PCR profile

166   was the same except that the annealing temperature was increased to 55 °C. Sequencing was

167   carried out using a BigDye197 Terminator Cycle Sequencing Ready Reaction Kit (Applied

168   Biosystems) and analysed with an ABI-3130XL DNA sequencer (Applied Biosystems).

169   Sequences were deposited in GenBank with the following accession numbers: MK823130-

170   MK823186. COI sequences were edited and aligned (ClustalW algorithm (Thompson,

171   Gibson, & Higgins, 2002)) using Bioedit (V.7.2.3 (Hall, 1999)). The final dataset including

172   published and new sequences consisted of 418 sequences from 61 sites (Supplementary Table

173   1).

174   *Lineage delineation*

175   We delimited genetic lineages following the approach described in Heger (2013). Briefly, to

176   obtain a general overview of the existing lineages, we first constructed the phylogenetic tree

177   based on the matrix of the unique sequences (haplotypes) among the 418 considered in this

178   study. The sequence lengths of the dataset vary from 430 to 620 bp (depending on the primers

179   used to barcode the isolated cells). We constructed both a Maximum Likelihood (ML) and a

180   Bayesian tree, with the RAxML algorithm (Stamatakis, Hoover, Rougemont, & Renner,

181   2008) and MrBayes v. 3.2.6 (Ronquist & Huelsenbeck, 2003), using in both cases a GTR + Γ

182   model. We then tested if the haplotypes were distributed randomly by comparing our

183   observed distribution with a null model obtained by haplotypes randomly attributed to

184   lineages (10000 replicates). The tree root was placed between two major clades (clade I

185   contains the lineage L, K, M, J and clade II contains the lineages C, DE, F, B, A, G, H, I) that

186   appeared well supported in earlier works (Heger et al., 2013; Kosakyan et al., 2012). Bi-

187   partition support values were evaluated with 1,000 bootstrap replicates. Bayesian MCMC

188   analysis was carried out with two simultaneous chains and 50,000,000 generations. Trees

189   were sampled every 1,000 generation and the burn-in was set at 25%. The trees were rooted

190   internally based on the topology of trees obtained in earlier works (Heger et al., 2013;

191   Kosakyan et al., 2016), which showed two major clades with maximum support; we rooted

192   the tree between these two clades. As both trees were congruent, we presented only the ML

193   tree and used the Bayesian analysis to evaluate the nodes posterior probabilities (pp). We used

194   three independent methods of lineage delimitation to compare our assignments with those of

195   Heger 2013: 1) Automatic Barcode Gap Discovery (ABGD) (Fontaneto, Flot, & Tang, 2015;

196   Puillandre, Lambert, Brouillet, & Achaz, 2011) using the ABGD web-server

197   http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html, 2) The sequence divergences using

198   the Kimura 2-parameter (Kimura, 1980; Nassonova, Smirnov, Fahrni, & Pawlowski, 2010)

199   using the "ape" package (V3.2 (Paradis, Claude, & Strimmer, 2004)) in R (V3.0.1. (R. Core

200   Team, 2014)), and 3) GMYC analysis performed with the SPLITS package, version 1.0-19

201   (Fujisawa & Barraclough, 2013) coded in R, version 3.1.2 (R. Core Team, 2014).

202   *Haplotype and lineage network*

203   Haplotypes (defined as genetic units separated by at least a single mutation) were assigned to

204   the previously determined lineages. Haplotype networks were constructed using minimum

205   spanning network analysis as implemented in the software PopART (V1.7 (Leigh & Bryant,

206   2015), Supplementary Table 2). Four main geographical zones (Eastern North America,

207   Western North America, Europe and Asia) were defined to highlight the distribution of the

208   haplotypes.

209   *Historical Biogeography*

210   We first tested whether the observed patterns could be due to chance or to a sampling bias by

211   calculating the observed beta diversity following (Legendre & De Cáceres, 2013). We then

212   tested if the observed beta diversity was higher than expected by chance. We simulated beta

213   diversity values under null expectations and compared them to observed beta diversity values

214   to obtain p-values and standardized effect sizes. Simulated beta diversity values were

215   calculated using the same approach (Legendre & De Cáceres, 2013) on a permuted site by

216   species matrix. Permutations were conducted using the permatswap algorithm of the R-

217   package "vegan" (Oksanen, Blanchet, & Kindt, 2015) which preserves column sums. This

218   allows us to randomly attribute species to station while preserving species total abundance.

219   In order to determinate whether species distribution areas were correlated with ecological

220   tolerance, we determined the climatic niche breadth for each species using the tolerance index

221   (Dolédec, Chessel, & Gimaret-Carpentier, 2000), with the R package "ade4" (Dray & Dufour,

222   2007). This index estimates niche breadth based on environmental tolerance (i.e. climate)

223   (Hurlbert, 1978; Thuiller, 2004) using the dispersion of geographic cells that contain the

224   target species in the climatic multivariate space. Low values of the index suggest narrow

225   tolerance while high values correspond to generalists. These indices were inferred based on

226   geographical coordinates for each occurrence (Supplementary Table 1) and interpolated

227   climate data sets (Bioclim, 19 variables) that were generated at 2.5 arcmin resolution from

228   meteorological data (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). We then estimated

229   distribution areas of the species based on the number of plots where a given species was

230   observed (i.e. a rough estimate of the spatial range) and plotted it against the estimated

231   climatic niche breadths.

232   To evaluate the evolutionary events that may explain the current distribution of *H. papilio*

233   (e.g. dispersal, extinction, range-switching, sympatry, vicariance and founder effect), we

234   estimated the ancestral distribution and the frequency of event counts in each of 1000

235   Biogeographic Stochastic Mapping (BSM) analyses, using the BioGeoBEARS package

236   (Matzke, 2013) in R (V3.0.1. (R. Core Team, 2014)). BioGeoBEARS allows for the

237   estimation of ancestral geographic ranges on dated phylogeny, comparing several models of

238   range evolution. We used the DEC model (Ree, Smith, & Baker, 2008) with two free

239   parameters: "d" (dispersal rate) and "e" (extinction rate), and a fixed cladogenetic model

240   (cladogenetic event allowed: vicariance, sympatric-subset speciation, and sympatric range-

241   copying). We also used a DEC model with an extra parameter, "j", which represents the

242   founder-event speciation, where the new species "jumps" to a range outside of the ancestral

243   range (DEC + j model). The comparison of these two models was performed using Akaike

244   Information Criterion (AIC). The age of the nodes of the rescaled BEAST Tree of *H. papilio*

245   was estimated by constraining its root to 7 Mya, which corresponds to the documented origin

246   of *Sphagnum* peatlands (Stenøien, Shaw, Shaw, Hassel, & Gunnarsson, 2010), as all known

247   lineages of *H. papilio* are restricted to these environments. This type of approach has been

248   used to identify the geographical origin of multicellular taxa of different ages, dispersion

249   strategies and lifestyles, including hyacinthoid monocots (Ali, Yu, Pfosser, & Wetschnig,

250   2012), chameleons (Tolley, Townsend, & Vences, 2013) and bees (Trunz, Packer, Vieu,

251   Arrigo, & Praz, 2016), and is used here, to our knowledge, for the first time in

252   microorganisms. This approach allowed us to test the two hypotheses: If the frequency of

253   event counts in each of 1000 BSM sustain low frequency of dispersal events, related to other

254 biogeographic events, the first hypothesis is supported (1.- Dispersal is low and/or slow

255 enough so that traces of glacial cycles are reflected in its extant diversity), and otherwise, the

256 second hypothesis is supported (2.- Dispersal is high and/or fast, and diversity would be

257 expected to be maximal where the largest expanses of *Sphagnum* peatlands are found today).

258

259 **Results**

260 *Lineage delineation and diversity*

261 The 418 mitochondrial COI sequences of *H. papilio* revealed the existence of 13 or 14 distinct

262 lineages (Fig. 2). The Kimura 2-parameter test suggested the existence of 14 lineages based

263 on a threshold of ≥1% sequence divergence. The Generalized Mixed Yule Coalescent

264 (GMYC) method yielded 13 lineages (lower and upper confidence intervals: 10 and 29

265 lineages, respectively; P = 0.046) based on single threshold methods. Finally, the Automatic

266 Barcode Gap Discovery (ABGD) method identified 13 lineages, using a distinctive barcoding

267 gap of 7%. One of these lineages, called here "M" has not been previously recorded. This

268 lineage was recovered from localities not included in previous studies (i.e. (Gomaa et al.,

269 2014; Heger et al., 2013; Kosakyan et al., 2012; Oliverio et al., 2015)). It was supported by all

270 analyses, although the Kimura 2-parameter test suggested dividing it into two (Fig. 2).

271

272 *Phylogenetic reconstruction*

273 Sequences from the previously overlooked lineage M diverged from all others (pp = 1) and

274 branched as a sister group to lineages J and K (Fig. 2). Only a single haplotype was retrieved

275 from Lineage E, and five from lineage D (*sensu* Heger et al. (2013)). Here again the genetic

276   divergence was low (i.e. at most six nucleotides difference between the sequence of lineage E,

277   and the five sequences of lineage D, all of which were separated by a single nucleotide (Fig.

278   3)).

279

280   *Haplotype network*

281   The haplotype network (Fig. 3) showed that some lineages (B, H, and L) were composed of

282   only a single haplotype, whereas others included several haplotypes, independently of the

283   number of individual cells barcoded. Some lineages were relatively rare (e.g. B, L and M with

284   seven, two, and seven individuals respectively) whilst others were extensively recorded (e.g.

285   lineage A was identified more than a hundred times). Null model analyses show that such a

286   pattern is not expected under random assembly of lineages (Supplementary Fig. 1).

287

288   *Spatial patterns of phylogenetic richness*

289   We found that the observed beta diversity was significantly higher than expected by chance

290   (SES = 1.71; p=0.99), showing a strong spatial structuring of diversity. We also found a

291   strong positive correlation between niche breadth as estimated using Dolédec tolerance

292   indices and distribution areas (R2 =0.75; P=0.001). We also found that lineages differed in

293   their climatic niches with some lineages preferring colder and drier conditions

294   (Supplementary Fig. 2, lineage H) and others preferring warmer conditions with abundant

295   precipitation (Supplementary Fig. 2, lineage I).

296   The geographical distribution of phylogenetic richness showed a clear contrast (Fig 1, 3).

297   Only four lineages (A, C, J and G) were recovered from all of Eurasia, five from Eastern

298 North America (A, F, K, J and M), and nine from Western North America (six in Alaska and

299 five in the Pacific Northwest, only two being shared between these two regions). Thus,

300 regional as well as overall diversity and diversity turnover were all higher in North America

301 than in Eurasia.

302 The distribution of the different lineages (Fig. 3) suggests that several haplotypes are specific

303 to certain geographical areas (B, DE, H, I and L occur only in Western North America, while

304 K and M occurred only in Eastern North America), whereas others were geographically

305 widespread (e.g. J is found throughout the Holarctic realm). Null model analyses show that

306 such a pattern is not expected under random assembly of lineages (Supplementary Fig. 1).

307 This structure in lineage distribution suggests that geographic dispersal has occurred

308 comparatively slowly, allowing it to be recovered with a genetic marker such as mtCOI used

309 for species-level delineation in this group of organisms (Kosakyan et al., 2012).

310

311 *Origin of lineages and evaluation of the diversification processes*

312 The AIC selection of biogeographic models implemented in BioGeoBEARS indicated that a

313 DEC model was the best-supported (Supplementary Table 3). Based on this model, the most

314 likely ancestral areas for *H. papilio* are in Western North America (Fig. 4). The dispersal

315 summary extracted from the 1,000 BSM's maps showed that most of the dispersal events

316 occurred from Western North America and Asia to the other biogeographic areas, and from

317 Asia to Europe (Supplementary Table 4). The results of the ancestral area estimation and

318 number of dispersal events analyses showed that the most frequent process during the

319 historical biogeography of *H. papilio* was narrow sympatry (i.e. when the ancestral range

320 contains one area, and both daughter lineages inherit that area), followed by a low frequency

321 of dispersal events (range expansion) (Supplementary Fig. 1). The importance of vicariance

322 and founder events were comparatively limited (Supplementary Fig. 1).

323 **Discussion**

324 *Diversity and geographical distribution of the lineages and haplotypes*

325 The species complex *H. papilio* is represented by at least 13 lineages in the Holarctic region,

326 one of which had not been previously described. Although it is possible that some lineages

327 remain to be discovered, our globally extensive sampling retrieved only one additional

328 lineage (M), suggesting that we have now captured most of the group's diversity. The genetic

329 distances determined by our taxon delineating approaches are consistent with the barcoding

330 gap (<4%) used to discriminate species in other related testate amoebae lineages (e.g. genus

331 *Nebela*, Hyalospheniidae). The above-mentioned lineages were defined as species under

332 multiple and independent concepts, including ecological, morphological and evolutionary

333 (Singer et al., 2018). This might imply that the lineages retrieved in the present study can all

334 be considered as separate species (Kosakyan, Gomaa, Mitchell, Heger, & Lara, 2013;

335 Kosakyan et al., 2012; Singer, Kosakyan, Pillonel, Mitchell, & Lara, 2015; Singer et al.,

336 2018). The accuracy of a species tree built on a single locus may be still questioned,

337 especially in the case of recent radiations, as the existence of several caveats (like sequencing

338 pseudogenes, ongoing hybridization processes) cannot be ruled out and may distort the tree's

339 topology. In Amoebozoa, COI has been chosen as the most accurate marker notably because

340 of its sensitivity and lack of intra-individual variability (Nassonova et al., 2010) and we

341 therefore consider it reliable.

342 Lineages of *H. papilio* show different distribution patterns over the Holarctic realm. Four

343 lineages (J, A, C, G) were found in several regions with contrasted climates (Fig. 1)

14

344   suggesting that they have a greater ecological tolerance. This is corroborated by the strong

345   correlation between climatic niche breadth and estimated distribution ranges (Fig. 5),

346   suggesting that colonization capacity is constrained by specific tolerance to climates. If these

347   distributional patterns reflect evolutionary adaptation to long-distance dispersal, it would then

348   imply that the required physiological/lifestyle adaptations to long range migration have

349   appeared independently at least four times in the history of the *H. papilio* species complex

350   (Fig. 4).

351   The existence of restricted distributions is even clearer at the haplotype level. Of the 74 total

352   *H. papilio* haplotypes, only seven (9.2%) were present in two zones, two in three zones

353   (2.6%) and no single haplotype was found in all four zones. This indicates that even

354   widespread lineages (e.g. lineage J) show high infra-specific genetic structuring, which

355   suggests limited gene flow among sites, and thus, geographical isolation (Fernández,

356   Hernández, Schiaffino, Izaguirre, & Lara, 2017; Lara, Heger, Scheihing, & Mitchell, 2010).

357   Hypothesis 2, that diversity is maximal where the largest expanses of *Sphagnum* peatlands are

358   found today cannot be supported by our data and analyses. Under this hypothesis, highest

359   diversity would be expected in regions such as Western Siberia where peatlands are at their

360   most extensive and cover more than 20% of the landscape (Peregon, Maksyutov, Kosykh, &

361   Mironycheva-Tokareva, 2008). However, only four "far travelled" lineages were found in the

362   entire of Eurasia, as compared to 13 in North America. This is despite larger overall area,

363   more extensive *Sphagnum* peatland extent and an extensive range of climatic conditions. Six

364   lineages, 50% more than the entire of Eurasia, were found only in Alaska. In contrast, most

365   genetic diversity seems to be located along the Western North American coast, a region where

366   peatlands are typically small and scattered today. This fact, together with the strong spatial

367 patterns in lineage distribution observed (Fig. 1, 4) advocates against our hypothesis 2 (fast

368 dispersal).

369 *Geographic origin and influence of historical events*

370 All Eurasian lineages identified were also present in North America, while several lineages

371 were restricted to North America. This observation alone suggests an American origin for *H.*

372 *papilio*. Our ancestral range reconstruction corroborates this inference, placing the most

373 probable origin of the *H. papilio* complex in Western North America (Fig. 4).

374 Dating speciation events is difficult in testate amoebae as their lineages cannot be

375 morphologically distinguished (Mulot et al., 2017); testate amoeba shell records in peat are

376 rare before the Holocene. Nevertheless, it is still possible to infer a time window for the

377 radiation of the lineages indirectly based on the very strict habitat specificity of this taxon. All

378 lineages of *H. papilio* thus-far identified are restricted to *Sphagnum* peatlands. It is therefore

379 reasonable to assume that this highly adapted taxon evolved within these ecosystems. The

380 oldest fossils of genus *Hyalosphenia* were described from the Triassic (*H. baueri* 220 MYA)

381 (Schönborn, Dörfelt, Foissner, Krienitz, & Schäffer, 1999). *H. baueri* shares some traits like

382 an "indistinctly vase shape" and the presence of an organic lip surrounding the aperture with

383 *H. papilio*. However, it is far from clear that both taxa are directly related. Firstly, it has been

384 shown that the genus *Hyalosphenia* is paraphyletic, as *H. papilio* and *H. elegans* are only

385 distantly related (Lahr et al., 2019; Lara, Heger, Ekelund, Lamentowicz, & Mitchell, 2008).

386 Furthermore, a rough calculation can rule out the possibility of a very old age for *H. papilio*.

387 "Standard" coxI (estimated for animals) mutation rates are typically in the range of a few

388 percent per million year (Ho & Lo, 2013; Papadopoulou, Anastasiou, & Vogler, 2010),

389 sometimes much higher (Ney, Frederick, & Schul, 2018). The most divergent *H. papilio*

390 sequences are separated by roughly 10%, which implies that, in order for the deepest

391 branching in the complex (see Fig. 4) to be 100 MYA old, the mutation rate would need to be

392 of 0.01% / MYA. This is far below all rates known to date, and even lower than the mutation

393 rate of cnidarians which are known for their extremely slow evolving mitochondria (Park et

394 al., 2012). By contrast, to obtain an age of 7 million years the mutation rate would need to be

395 1.3% / MYA, which is similar to the mutation rate of many arthropods and thus more

396 parsimonious than the alternative.

397 During the Pleistocene, large areas of North America were intermittently covered by ice

398 although ice-free refugia remained. The area of *Sphagnum* peatlands likely repeatedly

399 expanded during inter-glacial periods and contracted in response to glacial periods when ice

400 masses covered most of the landscape (Shaw et al., 2014). This period is also coetaneous with

401 most cladogenesis in the *H. papilio* phylogenetic history, which suggests a series of speciation

402 events by cladogenetic range evolution which may have occurred during interglacial periods

403 (Fig. 4). Indeed, our analyses show that at least 8 out of 12 cladogenesis events occurred

404 during the Pleistocene, immediately after the 2.5 MYA boundary (Fig. 4).

405 This hypothesis is also in line with the fact that the BioGeoBears analyses designated narrow

406 sympatry or the inheritance of the ancestral area of a range by both daughter lineages, as a key

407 process explaining the distribution of *H. papilio* lineages. At the onset of Quaternary

408 glaciations (2.58 Mya), one lineage probably existed in Eastern and two in Western North

409 America (Fig. 4). While the first lineage probably survived south of the ice sheet, where

410 conditions were wet enough to allow the development of peatlands (Shaw et al., 2010), the

411 two others were most likely confined to refugia in Western North America.

412 The location of these refugia is known to have shaped the distribution of plants (Eidesen et

413 al., 2013) and animals (Klütsch, Manseau, Anderson, Sinkins, & Wilson, 2017). In particular,

414 Eastern Beringia (today Alaska and Yukon Territory) was wet enough to support the growth

415 of *Sphagnum* mosses and *Sphagnum* peatlands (Shaw et al., 2013, 2014). These peatlands

416 allowed the survival of associated organisms, likely including the lineages of *H. papilio*. In

417 contrast, Western Beringia (today far eastern Russia) was too dry to support large expanses of

418 *Sphagnum* peatlands (Shaw et al., 2013, 2014) and likely constituted a barrier for the

419 migration of *H. papilio* westwards. Our data suggest that the colonization of the Palaearctic

420 region occurred recently, possibly after the last glaciation (Fig. 1). Western Siberia, which

421 was a cold desert during the Last Glacial Maximum became covered with peatlands after

422 11000 BP (Velichko, Timireva, Kremenetski, MacDonald, & Smith, 2011) and could have

423 constituted a bridge that facilitated the invasion of the Western Palaearctic by "far travelled"

424 lineages of *H. papilio*. Interestingly, a similar pattern has been suggested for the species

425 *Sphagnum angermanicum* (Stenøien et al., 2010).

426 The present-day distribution of lineages and the local palaeogeographical context designates

427 Eastern Beringia or the Pacific Coast as the most probable origin for all extant *H. papilio*

428 lineages. The higher diversity of *H. papilio* haplotypes in North America as compared to

429 Europe mirrors the higher diversity of vascular plants (Earl Latham & Ricklefs, 1993;

430 Svenning, 2003), and both were likely similarly driven by glaciations. The phylogeographic

431 history of *H. papilio*, used here as a convenient model taxon for protists lacking specialised

432 morphological adaptation for dispersal, thus highlights the importance of historical processes

433 in explaining the distribution of extant microbial diversity.

434 Thus, following a dispersal event, sympatric diversification could indeed have played a major

435 role in shaping the current phylogeography of *H. papilio* (Supplementary Fig. 1). It remains to

436   be determined if the case of *H. papilio* is representative for free-living microorganisms in

437   general. *H. papilio* is large by microbial standards; testate amoebae mostly range between 20

438   and 200μm and many other protists and most fungi and prokaryotes are smaller. *H. papilio* is

439   also restricted to *Sphagnum* mosses, which, although widespread across the Holarctic,

440   nevertheless constitute a very specific habitat. More generalist, smaller species and/or species

441   possessing structures adapted to dispersal (e.g. fruiting bodies as in many other Amoebozoa,

442   (Shadwick, Spiegel, Shadwick, Brown, & Silberman, 2009) may show patterns which agree

443   better with the second hypothesis. Elucidating the historical processes shaping the diversity of

444   protists with different dispersal strategies, and comparing patterns with better known

445   macroscopic organisms will open the way to understanding the processes of diversification

446   that produced the immense diversity existing today.

456

457

458

459

460 **References**

461 Ali, S. S., Yu, Y., Pfosser, M., & Wetschnig, W. (2012). Inferences of biogeographical histories within
462   subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in
463   RASP (Reconstruct Ancestral State in Phylogenies). *Annals of Botany*, *109*(1), 95–107. doi:
464   10.1093/aob/mcr274
465 Amesbury, M. J., Booth, R. K., Roland, T. P., Bunbury, J., Clifford, M. J., Charman, D. J., … Hughes, P. D.
466   M. (2018). Towards a Holarctic synthesis of peatland testate amoeba ecology: Development
467   of a new continental-scale palaeohydrological transfer function for North America and
468   comparison to European data. *Quaternary Science Reviews*, *201*, 483–500.
469 Amesbury, M. J., Swindles, G. T., Bobrov, A., Charman, D. J., Holden, J., Lamentowicz, M., … Payne, R.
470   J. (2016). Development of a new pan-European testate amoeba transfer function for
471   reconstructing peatland palaeohydrology. *Quaternary Science Reviews*, *152*, 132–151.
472 Arbogast, B. S., Drovetski, S. V., Curry, R. L., Boag, P. T., Seutin, G., Grant, P. R., … Anderson, D. J.
473   (2006). The origin and diversification of Galapagos mockingbirds. *Evolution*, *60*(2), 370–382.
474 Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P., & Slowinski, J. B. (2002). Estimating divergence
475   times from molecular data on phylogenetic and population genetic timescales. *Annual*
476   *Review of Ecology and Systematics*, *33*(1), 707–740.
477 Baas-Becking, L. G. M. (1934). *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon
478   NV.
479 Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., & Tedersoo, L. (2016). Stochastic
480   distribution of small soil eukaryotes resulting from high dispersal and drift in a local
481   environment. *The ISME Journal*, *10*(4), 885.
482 Dolédec, S., Chessel, D., & Gimaret-Carpentier, C. (2000). Niche separation in community analysis: a
483   new method. *Ecology*, *81*(10), 2914–2927.
484 Dray, S., & Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists.
485   *Journal of Statistical Software*, *22*(4), 1–20.
486 Duckert, C., Blandenier, Q., Kupferschmid, F. A. L., Kosakyan, A., Mitchell, E. A. D., Lara, E., & Singer,
487   D. (2018). En garde! Redefinition of *Nebela militaris* (Arcellinida, Hyalospheniidae) and
488   erection of *Alabasta* gen. nov. *European Journal of Protistology*, *66*, 156–165.
489 Earl Latham, R., & Ricklefs, R. E. (1993). Global patterns of tree species richness in moist forests:
490   energy-diversity theory does not account for variation in species richness. *Oikos*, *67*(2), 325–
491   333. doi: 10.2307/3545479
492 Eidesen, P. B., Ehrich, D., Bakkestuen, V., Alsos, I. G., Gilg, O., Taberlet, P., & Brochmann, C. (2013).
493   Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of
494   diversity. *New Phytologist*, *200*(3), 898–910. doi: 10.1111/nph.12412
495 Fenchel, T. (2005). Cosmopolitan microbes and their 'cryptic' species. *Aquatic Microbial Ecology*,
496   *41*(1), 49–54.
497 Fernández, L. D., Hernández, C. E., Schiaffino, M. R., Izaguirre, I., & Lara, E. (2017). Geographical
498   distance and local environmental conditions drive the genetic population structure of a
499   freshwater microalga (Bathycoccaceae; Chlorophyta) in Patagonian lakes. *FEMS Microbiology*
500   *Ecology*, *93*(10). doi: 10.1093/femsec/fix125
501 Fernández, L. D., Lara, E., & Mitchell, E. A. D. (2015). Checklist, diversity and distribution of testate
502   amoebae in Chile. *European Journal of Protistology*, *51*(5), 409–424.
503 Finlay, B. J. (1998). The global diversity of protozoa and other small species. *International Journal for*
504   *Parasitology*, *28*(1), 29–48.
505 Foissner, W. (1999). Protist diversity: estimates of the near-imponderable. *Protist*, *150*(4), 363–368.
506 Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of
507   mitochondrial cytochrome c oxidase subunit 1 from diverse metazoan invertebrates.

Molecular Marine Biology and Biotechnology, (5). Retrieved from
http://www.vliz.be/en/imis?refid=64543

Fontaneto, D., Flot, J.-F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the
meiofauna. *Marine Biodiversity*, *45*(3), 433–451. doi: 10.1007/s12526-015-0319-7

Forest, F., Colville, J. F., & Cowling, R. M. (2018). Evolutionary diversity patterns in the Cape flora of
South Africa. In *Phylogenetic Diversity* (pp. 167–187). Springer.

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the
generalized mixed yule coalescent approach: a revised method and evaluation on simulated
data sets. *Systematic Biology*, *62*(5), 707–724. doi: 10.1093/sysbio/syt033

Geisen, S., Fiore-Donno, A. M., Walochnik, J., & Bonkowski, M. (2014). *Acanthamoeba* everywhere:
high diversity of *Acanthamoeba* in soils. *Parasitology Research*, *113*(9), 3151–3158.

Geisen, S., Mitchell, E. A. D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., … Lara, E. (2018). Soil
protists: a fertile frontier in soil biology research. *FEMS Microbiology Reviews*, *42*(3), 293–
323. doi: 10.1093/femsre/fuy006

Gomaa, F., Kosakyan, A., Heger, T. J., Corsaro, D., Mitchell, E. A. D., & Lara, E. (2014). One alga to rule
them all: unrelated mixotrophic testate amoebae (Amoebozoa, Rhizaria and Stramenopiles)
share the same symbiont (Trebouxiophyceae). *Protist*, *165*(2), 161–176. doi:
10.1016/j.protis.2014.01.002

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

Heger, T. J., Mitchell, E. A. D., & Leander, B. S. (2013). Holarctic phylogeography of the testate
amoeba *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity
explained more by environment than dispersal limitation. *Molecular Ecology*, *22*(20), 5172–
5184. doi: 10.1111/mec.12449

Herbert, T. D., Lawrence, K. T., Tzanova, A., Cleaveland Peterson, L., Caballero-Gill, R., & Kelly, C. S.
(2016). Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience*,
9(11), 843–847. doi: 10.1038/ngeo2813

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution
interpolated climate surfaces for global land areas. *International Journal of Climatology*,
25(15), 1965–1978.

Ho, S. Y. W., & Lo, N. (2013). The insect molecular clock. *Australian Journal of Entomology*, *52*(2),
101–105.

Hurlbert, S. H. (1978). The measurement of niche overlap and some relatives. *Ecology*, *59*(1), 67–77.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through
comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–
120.

Klütsch, C. F. C., Manseau, M., Anderson, M., Sinkins, P., & Wilson, P. J. (2017). Evolutionary
reconstruction supports the presence of a Pleistocene Arctic refugium for a large mammal
species. *Journal of Biogeography*, *44*(12), 2729–2739. doi: 10.1111/jbi.13090

Kosakyan, A., Gomaa, F., Mitchell, E. A. D., Heger, T. J., & Lara, E. (2013). Using DNA-barcoding for
sorting out protist species complexes: A case study of the *Nebela tincta–collaris–bohemica*
group (Amoebozoa; Arcellinida, Hyalospheniidae). *European Journal of Protistology*, *49*(2),
222–237. doi: 10.1016/j.ejop.2012.08.006

Kosakyan, A., Heger, T. J., Leander, B. S., Todorov, M., Mitchell, E. A. D., & Lara, E. (2012). COI
barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity
and redefinition of the Hyalospheniidae Schultze. *Protist*, *163*(3), 415–434. doi:
10.1016/j.protis.2011.10.003

Kosakyan, A., Lahr, D. J. G., Mulot, M., Meisterfeld, R., Mitchell, E. A. D., & Lara, E. (2016).
Phylogenetic reconstruction based on COI reshuffles the taxonomy of hyalosphenid shelled

557   (testate) amoebae and reveals the convoluted evolution of shell plate shapes. *Cladistics*,
558   *32*(6), 606–623. doi: 10.1111/cla.12167

559 Kostka, J. E., Weston, D. J., Glass, J. B., Lilleskov, E. A., Shaw, A. J., & Turetsky, M. R. (2016). The
560   *Sphagnum* microbiome: new insights from an ancient plant lineage. *New Phytologist*, *211*(1),
561   57–64.

562 Lahr, D. J. G., Kosakyan, A., Lara, E., Mitchell, E. A. D., Morais, L., Porfirio-Sousa, A. L., … Kang, S.
563   (2019). Phylogenomics and morphological reconstruction of Arcellinida testate amoebae
564   highlight diversity of microbial eukaryotes in the Neoproterozoic. *Current Biology*.

565 Lara, E., Heger, T. J., Ekelund, F., Lamentowicz, M., & Mitchell, E. A. D. (2008). Ribosomal RNA genes
566   challenge the monophyly of the Hyalospheniidae (Amoebozoa: Arcellinida). *Protist*, *159*(2),
567   165–176.

568 Lara, E., Heger, T. J., Scheihing, R., & Mitchell, E. A. D. (2010). COI gene and ecological data suggest
569   size-dependent high dispersal and low intra-specific diversity in free-living terrestrial protists
570   (Euglyphida: *Assulina*). *Journal of Biogeography*, *38*(4), 640–650. doi: 10.1111/j.1365-
571   2699.2010.02426.x

572 Lara, E., Heger, T. J., Scheihing, R., & Mitchell, E. A. D. (2011). COI gene and ecological data suggest
573   size-dependent high dispersal and low intra-specific diversity in free-living terrestrial protists
574   (Euglyphida: *Assulina*). *Journal of Biogeography*, *38*(4), 640–650.

575 Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: dissimilarity
576   coefficients and partitioning. *Ecology Letters*, *16*(8), 951–963.

577 Leigh, J. W., & Bryant, D. (2015). popart: full-feature software for haplotype network construction.
578   *Methods in Ecology and Evolution*, *6*(9), 1110–1116. doi: 10.1111/2041-210X.12410

579 Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., … Kuske,
580   C. R. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews*
581   *Microbiology*, *4*(2), 102.

582 Matzke, N. J. (2013). *Probabilistic historical niogeography: new models for founder-event speciation,*
583   *imperfect detection, and fossils allow improved accuracy and model-testing* (UC Berkeley).
584   Retrieved from https://escholarship.org/uc/item/8227p52c

585 Mulot, M., Marcisz, K., Grandgirard, L., Lara, E., Kosakyan, A., Robroek, B. J. M., … Mitchell, E. A. D.
586   (2017). Genetic determinism vs. phenotypic plasticity in protist morphology. *Journal of*
587   *Eukaryotic Microbiology*, *64*(6), 729–739. doi: 10.1111/jeu.12406

588 Mutinová, P. T., Neustupa, J., Bevilacqua, S., & Terlizzi, A. (2016). Host specificity of epiphytic diatom
589   (Bacillariophyceae) and desmid (Desmidiales) communities. *Aquatic Ecology*, *50*(4), 697–709.

590 Nassonova, E., Smirnov, A., Fahrni, J., & Pawlowski, J. (2010). Barcoding amoebae: comparison of
591   SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae.
592   *Protist*, *161*(1), 102–115. doi: 10.1016/j.protis.2009.07.003

593 Ney, G., Frederick, K., & Schul, J. (2018). A Post-pleistocene Calibrated Mutation Rate from Insect
594   Museum Specimens. *PLoS Currents*, *10*.

595 Oksanen, J., Blanchet, F. G., & Kindt, R. (2015). *Vegan: commity ecology package. R package version*
596   *2.3-0*.

597 Oliverio, A. M., Lahr, D. J. G., Grant, J., & Katz, L. A. (2015). Are microbes fundamentally different than
598   macroorganisms? Convergence and a possible case for neutral phenotypic evolution in
599   testate amoeba (Amoebozoa: Arcellinida). *Royal Society Open Science*, *2*(12), 150414. doi:
600   10.1098/rsos.150414

601 Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular
602   clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27*(7), 1659–1672.

603 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R
604   language. *Bioinformatics*, *20*(2), 289–290. doi: 10.1093/bioinformatics/btg412

605 Park, E., Hwang, D.-S., Lee, J.-S., Song, J.-I., Seo, T.-K., & Won, Y.-J. (2012). Estimation of divergence
606   times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil
607   record. *Molecular Phylogenetics and Evolution*, *62*(1), 329–345.

608 Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., … de Vargas, C. (2012). CBOL protist
609   working group: barcoding eukaryotic richness beyond the animal, plant, and fungal
610   kingdoms. *PLOS Biology*, *10*(11), e1001419. doi: 10.1371/journal.pbio.1001419

611 Peregon, A., Maksyutov, S., Kosykh, N. P., & Mironycheva-Tokareva, N. P. (2008). Map-based
612   inventory of wetland biomass and net primary production in western Siberia. *Journal of
613   Geophysical Research: Biogeosciences*, *113*(G1).

614 Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2011). ABGD, Automatic Barcode Gap Discovery
615   for primary species delimitation. *Molecular Ecology*, *21*(8), 1864–1877. doi: 10.1111/j.1365-
616   294X.2011.05239.x

617 R. Core Team. (2014). *R: A language and environment for statistical computing. R Foundation for
618   Statistical Computing, Vienna, Austria. 2013*. ISBN 3-900051-07-0.

619 Ree, R. H., Smith, S. A., & Baker, A. (2008). Maximum likelihood inference of geographic range
620   evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*(1), 4–14.
621   doi: 10.1080/10635150701883881

622 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed
623   models. *Bioinformatics*, *19*(12), 1572–1574. doi: 10.1093/bioinformatics/btg180

624 Ryšánek, D., Hrčková, K., & Škaloud, P. (2015). Global ubiquity and local endemism of free-living
625   terrestrial protists: phylogeographic assessment of the streptophyte alga *Klebsormidium*.
626   *Environmental Microbiology*, *17*(3), 689–698.

627 Santoferrara, L. F., Rubin, E., & Mcmanus, G. B. (2018). Global and local DNA (meta) barcoding reveal
628   new biogeography patterns in tintinnid ciliates. *Journal of Plankton Research*, *40*(3), 209–
629   221.

630 Schönborn, W., Dörfelt, H., Foissner, W., Krienitz, L., & Schäffer, U. (1999). A fossilized microcenosis
631   in Triassic amber. *Journal of Eukaryotic Microbiology*, *46*(6), 571–584.

632 Schönswetter, P., Stehlik, I., Holderegger, R., & Tribsch, A. (2005). Molecular evidence for glacial
633   refugia of mountain plants in the European Alps. *Molecular Ecology*, *14*(11), 3547–3555.

634 Shadwick, L. L., Spiegel, F. W., Shadwick, J. D. L., Brown, M. W., & Silberman, J. D. (2009).
635   *Eumycetozoa= Amoebozoa*?: SSUrDNA phylogeny of protosteloid slime molds and its
636   significance for the amoebozoan supergroup. *PloS One*, *4*(8), e6754.

637 Shaw, A. J., Carter, B. E., Aguero, B., da Costa, D. P., & Crowl, A. A. (2019). Range change evolution of
638   peat mosses (*Sphagnum*) within and between climate zones. *Global Change Biology*, *25*(1),
639   108–120.

640 Shaw, A. J., Devos, N., Cox, C. J., Boles, S. B., Shaw, B., Buchanan, A. M., … Seppelt, R. (2010).
641   Peatmoss (*Sphagnum*) diversification associated with Miocene Northern Hemisphere climatic
642   cooling? *Molecular Phylogenetics and Evolution*, *55*(3), 1139–1145. doi:
643   10.1016/j.ympev.2010.01.020

644 Shaw, A. J., Golinski, G. K., Clark, E. G., Shaw, B., Stenøien, H. K., & Flatberg, K. I. (2013).
645   Intercontinental genetic structure in the amphi-Pacific peatmoss *Sphagnum miyabeanum*
646   (Bryophyta: Sphagnaceae). *Biological Journal of the Linnean Society*, *111*(1), 17–37. doi:
647   10.1111/bij.12200

648 Shaw, A. J., Shaw, B., Stenøien, H. K., Golinski, G. K., Hassel, K., & Flatberg, K. I. (2014). Pleistocene
649   survival, regional genetic structure and interspecific gene flow among three northern peat-
650   mosses: *Sphagnum inexspectatum*, *S. orientale* and *S. miyabeanum*. *Journal of Biogeography*,
651   *42*(2), 364–376. doi: 10.1111/jbi.12399

652 Singer, D., Kosakyan, A., Pillonel, A., Mitchell, E. A. D., & Lara, E. (2015). Eight species in the *Nebela
653   collaris* complex: *Nebela gimlii* (Arcellinida, Hyalospheniidae), a new species described from a

654 Swiss raised bog. *European Journal of Protistology*, *51*(1), 79–85. doi:
655 10.1016/j.ejop.2014.11.004

656 Singer, D., Kosakyan, A., Seppey, C. V. W., Pillonel, A., Fernández, L. D., Fontaneto, D., … Lara, E.
657 (2018). Environmental filtering and phylogenetic clustering correlate with the distribution
658 patterns of cryptic protist species. *Ecology*, *99*(4), 904–914. doi: 10.1002/ecy.2161

659 Škaloud, P., Škaloudová, M., Doskočilová, P., Kim, J. Im., Shin, W., & Dvořák, P. (2019). Speciation in
660 protists: Spatial and ecological divergence processes cause rapid species diversification in a
661 freshwater chrysophyte. *Molecular Ecology*.

662 Šlapeta, J., López-García, P., & Moreira, D. (2005). Global dispersal and ancient cryptic species in the
663 smallest marine eukaryotes. *Molecular Biology and Evolution*, *23*(1), 23–29.

664 Smith, H. G., Bobrov, A., & Lara, E. (2007). Diversity and biogeography of testate amoebae. In *Protist*
665 *Diversity and Geographical Distribution* (pp. 95–109). Springer.

666 Stamatakis, A., Hoover, P., Rougemont, J., & Renner, S. (2008). A rapid bootstrap algorithm for the
667 RAxML web servers. *Systematic Biology*, *57*(5), 758–771. doi: 10.1080/10635150802429642

668 Stenøien, H. K., Shaw, A. J., Shaw, B., Hassel, K., & Gunnarsson, U. (2010). North American origin and
669 recent European establishments of the Amphi-Atlantic peat moss *Sphagnum Angermanicum*.
670 *Evolution*, *65*(4), 1181–1194. doi: 10.1111/j.1558-5646.2010.01191.x

671 Svenning, J.-C. (2003). Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree
672 flora. *Ecology Letters*, *6*(7), 646–653. doi: 10.1046/j.1461-0248.2003.00477.x

673 Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW
674 and ClustalX. *Current Protocols in Bioinformatics*, 2.3.1-2.3.22.

675 Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global*
676 *Change Biology*, *10*(12), 2020–2027.

677 Tolley, K. A., Townsend, T. M., & Vences, M. (2013). Large-scale phylogeny of chameleons suggests
678 African origins and Eocene diversification. *Proceedings of the Royal Society of London B:*
679 *Biological Sciences*, *280*(1759), 20130184. doi: 10.1098/rspb.2013.0184

680 Treat, C. C., Kleinen, T., Broothaerts, N., Dalton, A. S., Dommain, R., Douglas, T. A., … Brovkin, V.
681 (2019). Widespread global peatland establishment and persistence over the last 130,000 y.
682 *Proceedings of the National Academy of Sciences*, *116*(11), 4822. doi:
683 10.1073/pnas.1813305116

684 Trunz, V., Packer, L., Vieu, J., Arrigo, N., & Praz, C. J. (2016). Comprehensive phylogeny, biogeography
685 and new classification of the diverse bee tribe Megachilini: Can we use DNA barcodes in
686 phylogenies of large genera? *Molecular Phylogenetics and Evolution*, *103*, 245–259. doi:
687 10.1016/j.ympev.2016.07.004

688 van Breemen, N. (1995). How *Sphagnum* bogs down other plants. *Trends in Ecology & Evolution*,
689 *10*(7), 270–275. doi: 10.1016/0169-5347(95)90007-1

690 Velichko, A. A., Timireva, S. N., Kremenetski, K. V., MacDonald, G. M., & Smith, L. C. (2011). West
691 Siberian Plain as a late glacial desert. *Quaternary International*, *237*(1–2), 45–53.

692 Wilkinson, D. M. (2001). What is the upper size limit for cosmopolitan distribution in free-living
693 microorganisms? *Journal of Biogeography*, *28*(3), 285–291.

694 Wilkinson, D. M., Koumoutsaris, S., Mitchell, E. A. D., & Bey, I. (2011). Modelling the effect of size on
695 the aerial dispersal of microorganisms. *Journal of Biogeography*, *39*(1), 89–97. doi:
696 10.1111/j.1365-2699.2011.02569.x

697
698

699

Data Accessibility

The 57 DNA sequences of COI gene of *Hyalophenia papilio* are available in in GenBank with the following accession numbers: MK823130- MK823186.

Author contributions

D.S., E.A.D.M. and E.L. designed the experiments; D.S., E.A.D.M., G.G., H.R., L.B., N.G.K, I.G., L.I.H., K.K., M.L., N.P.K., R.J.P. and K.V., collected the samples; D.S., Q.B., C.D., L.D.F., B.F., C.E.H. and E.L. analysed the data; D.S., E.A.D.M., R.J.P. and E.L. wrote the first version of the manuscript, which was then edited by all co-authors.

Conflict of interest

The material in this manuscript is original research, has not been previously published and has not been submitted for publication elsewhere while under consideration for Molecular Ecology. We would also like that only the online version appears in colour, as we took special care in building our figures in a way that they can be read in black and white as well. We have no conflict of interest in this research.
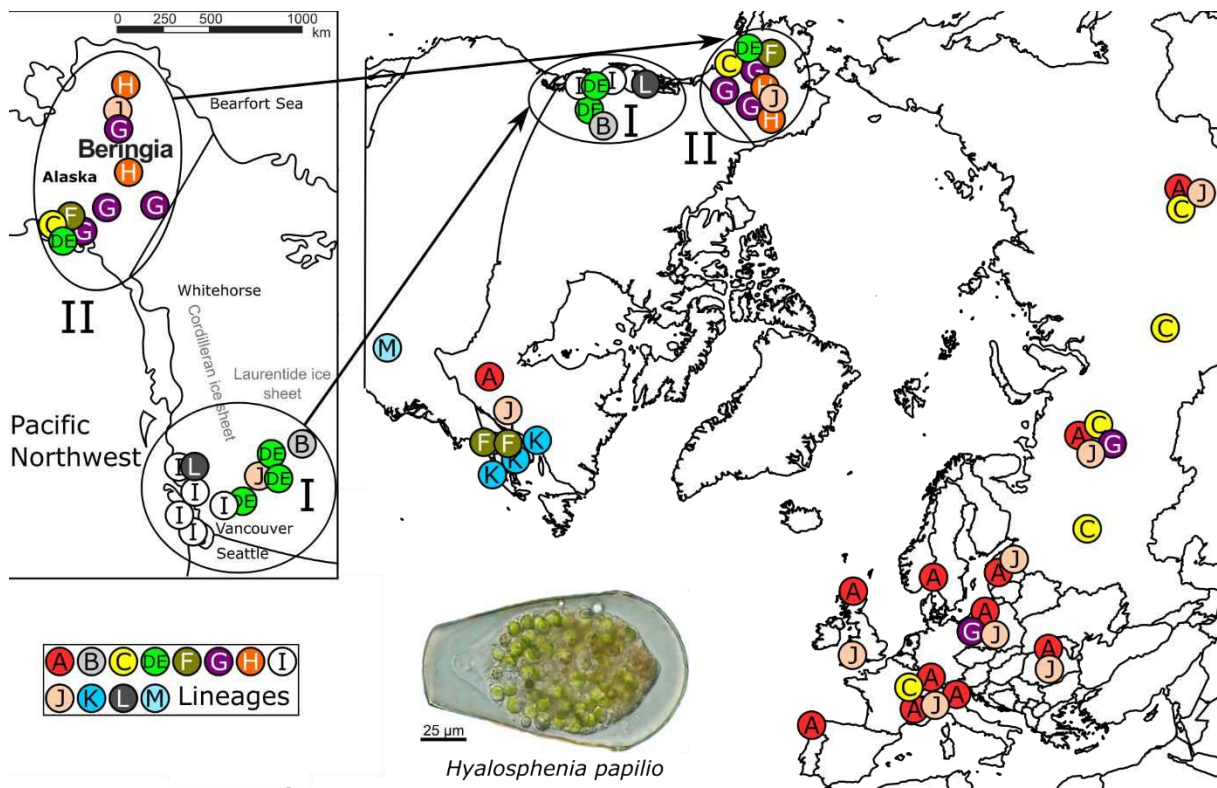
722

724 Fig. 1) Holarctic distribution of *Hyalosphenia papilio* lineages. Each circle corresponds to a
725 sampling site where the lineage has been detected. Lineage codes correspond to phylogenetic
726 groups, as identified in Heger et al. 2013. I and II present a detailed representation of Beringia
727 area. Inset: Light microscopy image of *Hyalosphenia papilio*. The pyriform outline
728 corresponds to the shell that protects the single-cell body of the organism and its
729 endosymbiotic microalgae (green dots)

730
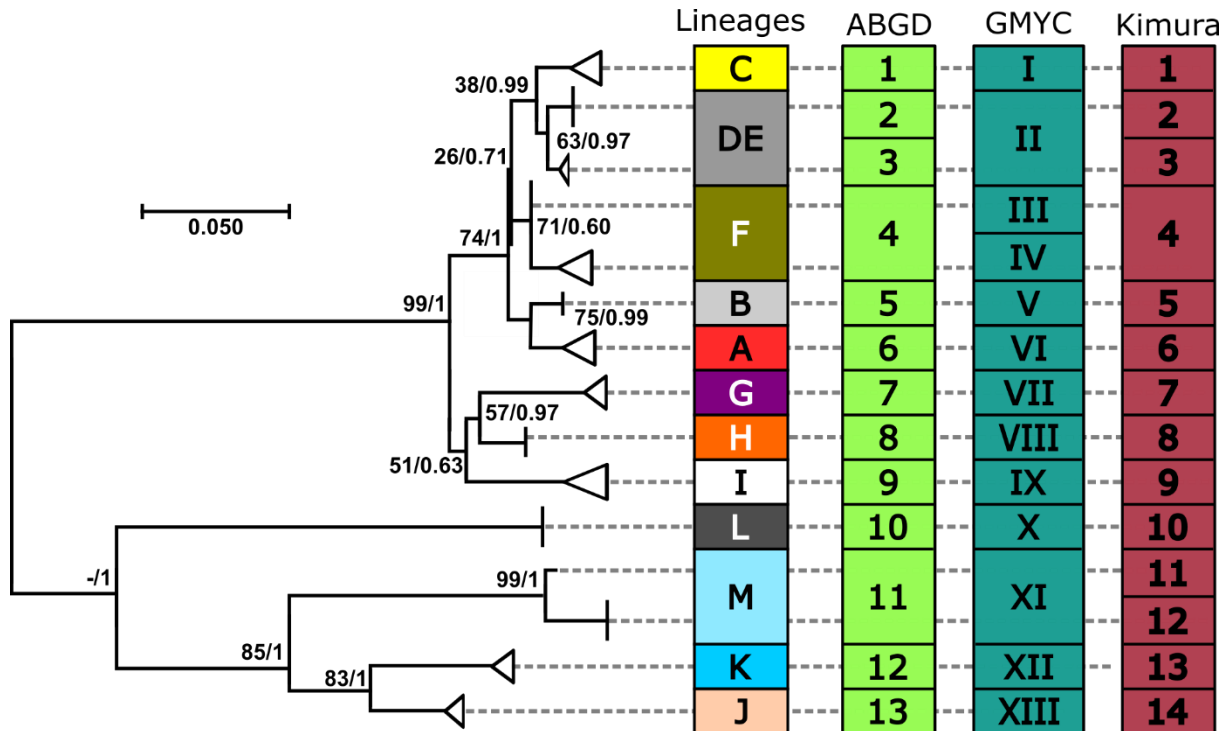
731

732 Fig. 2) Maximum Likelihood (ML) and Bayesian concatenated phylogenetic tree from 76
733 unique sequences of *Hyalosphenia papilio* isolated from *Sphagnum* peatlands across the
734 Holarctic realm. Numbers along branches represent, respectively, bootstraps obtained by ML
735 and posterior probabilities as calculated with Bayesian analyses. Trees were rooted internally
736 based on the topology of trees obtained in earlier works (Kosakyan et al. 2016, Heger et al.
737 2013), which showed two major clades with maximum support; we rooted the tree between
738 these two clades. The tree also represents the different lineages obtained with the ABGD and
739 GMYC analyses and the Kimura 2-parameter test.

740

27
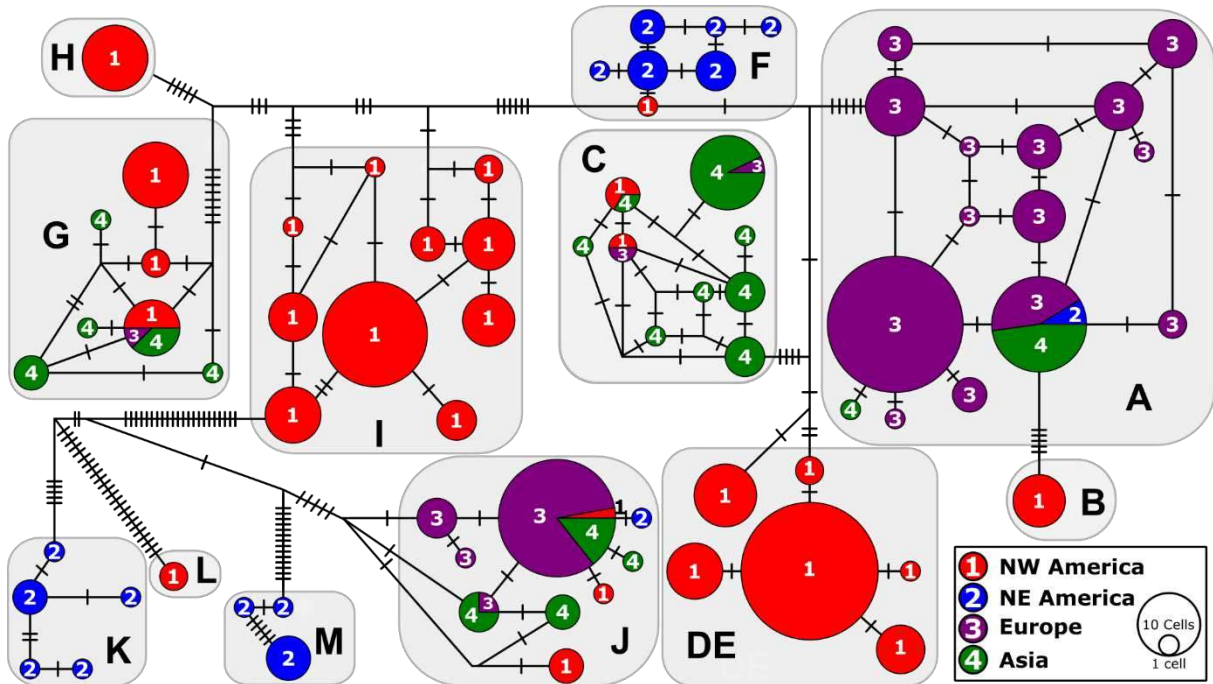
Fig. 3) Median joining haplotype network of cytochrome oxidase subunit 1 (COI) gene of *Hyalosphenia papilio* from *Sphagnum* peatlands in the Holarctic realm. Grey boxes and letters represent the different lineages identified in the present study. Colours indicate geographical regions (legend: bottom right inset). Circle sizes are proportional to the number of sequenced single cells of *H. papilio* within each haplotype. Cross lines show the number of mutational steps between haplotypes.

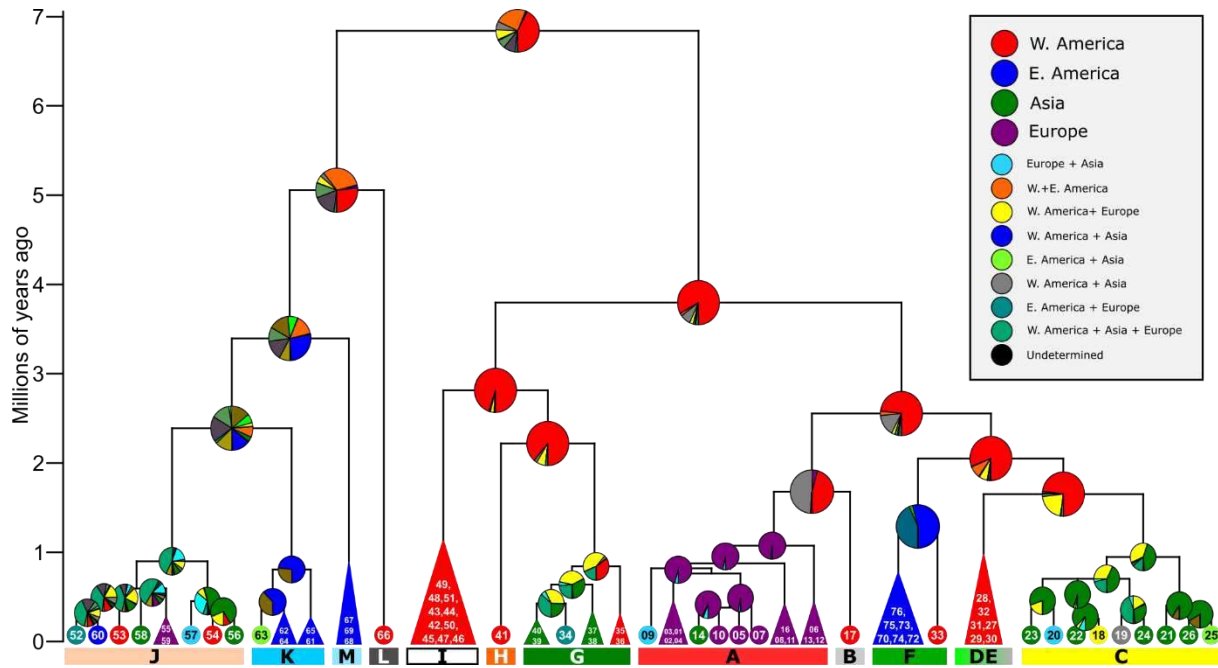Fig. 4) Biogeographical analysis of *H. papilio* from *Sphagnum* peatlands across the Holarctic realm using BioGeoBEARS. The four biogeographical areas are: Eastern North America (in blue), Western North America (in red), Europe (in purple) and Asia (in green). Pie charts at nodes indicate support for each area. The tips are labelled with present-day distributions. The secondary colours indicate range combinations of the tip ranges.
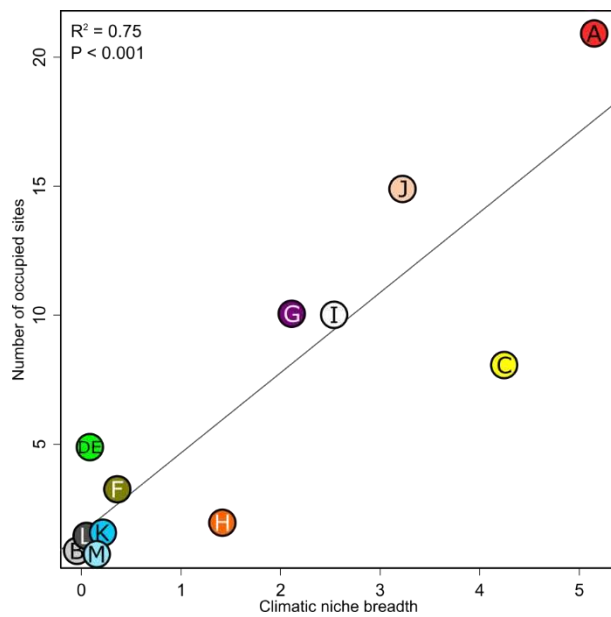
Fig. 5) Relationship of lineage climatic niche breadth to lineage range size (estimated as the number of locations where a given lineage is present) of Hyalosphenia papilio. Climatic niche breadth was estimating from Bioclim variables using the tolerance index of Doledec et al. (2000)

29