Phylogenetic and phylogeographic patterns of *Didymosphenia geminata* on invaded sites in Chile

Bernardita Cayupe [Corresp., 1] , **Nicole Ehrenfeld** [1] , **Rodrigo Moreno** [2] , **Fabio A. Labra** [2,3] , **Carolina Díaz** [4]

[1] Centro de Investigación AustralBiotech, Universidad Santo Tomás, Santiago, Chile

[2] Facultad de Ciencias, Universidad Santo Tomás, Santiago, Chile

[3] Centro de Investigación e Innovación para el Cambio Climático, Universidad Santo Tomás, Santiago, Chile

[4] Amakaik Consultoría Ambiental SpA, Santiago, Chile

Corresponding Author: Bernardita Cayupe
Email address: bcayupe@australbiotech.cl

Understanding the population dynamics of invasive processes has become a pressing concern in a highly connected world faced with ongoing climate change and increased exotic species introduction. In Chile, the invasive freshwater benthic diatom *Didymosphenia geminata* (Lyngbye) Schmidt has become widespread, expanding across multiple river basins spanning over 3000 km in three years. Here we evaluate the phylogenetic and phylogeographic relationships of *D. geminata* samples collected throughout the invaded range in Chile, using specific rbcL gene sequence previously published. Genetic sequences for this marker were generated for 19 sample sites, and were compared with available freshwater diatom sequences, as well as with previously published rbcL gene sequences for *D. geminata*. We find that all genetic sequences collected within Chile present phylogenetic divergences from *D. geminata* samples collected in Siberia, as well as from samples of the genera Gomphonema, Cymbella and Encyonema. Thus, we validate the invasion by *D. geminata*, in agreement with existing morphological taxonomic criteria. In addition, a haplotype analysis showed a total of 13 haplotypes, two of which (halpotypes I and IX) found in 12 and 3 populations respectively, while each of the remaining haplotypes presents a single population. Thus, these results are consistent either with the introduction of multiple lineages, or with a rapid genetic differentiation in this invading freshwater diatom. Further genetic sampling, both within Chile and in countries that may have been potential sources of the invasion are needed to test these alternative hypotheses.

Phylogenetic and phylogeographic patterns of *Didymosphenia geminata* on invaded sites in Chile

Bernardita Cayupe*,1 Nicole Ehrenfeld1, Rodrigo Moreno2, Fabio A. Labra2,3, Carolina Diaz4 .

1 Austral Biotech, Universidad Santo Tomás, Santiago, Chile.
2 Facultad de Ciencias, Universidad Santo Tomás, Santiago, Chile.
3 Centro de Investigación e Innovación para el Cambio Climático, Universidad Santo Tomás, Santiago, Chile.
4Amakaik Consultoría Ambiental SpA, Santiago, Chile.

Corresponding author: *E –mail address: bcayupe@australbiotech.cl

Abstract

Understanding the population dynamics of invasive processes has become a pressing concern in a highly connected world faced with ongoing climate change and increased exotic species introduction. In Chile, the invasive freshwater benthic diatom *Didymosphenia geminata (Lyngbye) Schmidt* has become widespread, expanding across multiple river basins spanning over 3000 km in three years. Here we evaluate the phylogenetic and phylogeographic relationships of *D. geminata* samples collected throughout the invaded range in Chile, using specific *rbc*L gene sequence previously published. Genetic sequences for this marker were generated for 19 sample sites, and were compared with available freshwater diatom sequences, as well as with previously published *rbc*L gene sequences for *D. geminata*. We find that all genetic sequences collected within Chile present phylogenetic divergences from *D. geminata* samples collected in Siberia, as well as from samples of the genera *Gomphonema*, *Cymbella* and *Encyonema*. Thus, we validate the invasion by *D. geminata*, in agreement with existing morphological taxonomic criteria. In addition, a haplotype analysis showed a total of 13 haplotypes, two of which (halpotypes I and IX) found in 12 and 3 populations respectively, while each of the remaining haplotypes presents a single population. Thus, these results are consistent either with the introduction of multiple lineages, or with a rapid genetic differentiation in this invading freshwater diatom. Further genetic sampling, both within Chile and in countries that may have been potential sources of the invasion are needed to test these alternative hypotheses.

Introduction

Didymosphenia geminata (Lyngbye) Schmidt is a freshwater benthic diatom, also known as “Rock snot” because it may attach to solid surfaces with polysaccharide stalks extruded from individual cells, forming a dense mucilaginous mat which may extend to form nuisance blooms that dominate riverine environments (Bothwell et al., 2009). *D. geminata* is considered native to the northern hemisphere (Kilroy, 2004), with the first nuisance bloom documented on Vancouver Island, British Columbia in the early 1990s (Bothwell et al., 2009). In Chile, *D. geminata* was first reported in 1962 (Díaz et al., 2012; AMAKAIK, 2016). However, no further records were described in the literature until the summer of 2010 in the Futaleufú basin, were the first nuisance bloom was reported in Río Espolón (approximately 43° S – 71°O), Región de Los Lagos (Díaz et al., 2012; AMAKAIK, 2016; Segura 2011). Within three years, *D. geminata* expanded its range to 3000 km, ranging from 38° S to 53° S and it is now considered a ‘plague’, as declared by the Chilean Subsecretaria de Pesca, Resolución Ex. 2064/2010 (Leone et al., 2014).

The most common way to identify *D. geminata* in Chile has been the taxonomic evaluation under microscope (Díaz et al., 2012; AMAKAIK, 2016) However, recent work by Jaramillo *et al.* (2015) applied DNA-barcoding based on *rbc*L and 18s rRNA genes to identify current genetic lineages of *D. geminata* found in 4 Chilean rivers, thereby providing evidence that genetic identification of this invasive diatom is feasible. Nevertheless, available genetic sequences for species of the *Didymosphenia* genus still remain scarce in reference databases such as GenBank (see Nakov et al. 2014, Jaramillo et al. 2015). Hence, complementing available

genetic information could improve the information available to allow the early detection and monitoring of this invasive species particularly in uncontaminated rivers, helping managers to make conservation decisions about places invaded by *D. geminata* (Bothwell et al., 2009; Darling and Blum, 2007). In order to provide additional genetic evidence that allows us to examine the pattern of genetic diversity in this invasive species, we reconstructed intraspecific phylogeographic relationships and analyzed the patterns of haplotype frequencies in this species. To this end, we used specific ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbc*L) gene primers designed by Jaramillo et al., (2015) and examined the genetic sequences of *D. geminata* samples from 19 infected sites in Chile.

Considering that molecular markers can provide valuable information about biodiversity, relationships between environmental and may also allow the reconstruction of changes in biodiversity through time, the goal of this paper was to evaluate phylogenetic and phylogeographic relationships of *D. geminata* samples collected throughout the invaded range in Chile

Materials and methods

To evaluate phylogenetic relationships between sites infected with *D. geminata,* samples were collected between December of 2015 and January of 2016 from 19 sampling sites located from 38°S to 47°S in Central-Southern Chile (Table 1). At each sampling site, at least 20 mL of biomass were taken from mats produced by

D. geminata. Genomic samples were conserved with ethanol 70% in 50mL conical plastic tubes and were cold stored until they were moved to Austral Biotech Investigation Center, usually between 2 days. Once arrived to the laboratory dependences, were stored at -80°C until processing. Also, at each sampling site a 1 cm3 volume of *D. geminata* mat was collected and fixed using Lugol as described in Díaz et al. (2012) and AMAKAIK (2016), and then transported to laboratory to carry out taxonomic confirmation under light microscope. Biosecurity procedures were conducted at each sampling site to prevent potential *D. geminata* contamination among rivers, following the procedures described by Diaz et al. (2011).

To extract total genomic DNA (gDNA) from mat samples, Powersoil® kit from MoBio Laboratories were used. The quality and quantity of genomic DNA was evaluated using agarose gel electrophoresis and Optizen POP Spectrophotometer (Mecasys). To amplify the specific sequence of *rbc*L gene from *D. geminta*, we used the primers designed by Jaramillo et al. (2015), with the following sequences: rbcl-F: 5'-ACC AAC AAC TGT ACC AGC GT-3' and rbcl-R: 5'-TGG GAT GCT TCA TAC GCA GT-3'.

All PCRs were performed in a final volume of 20 µL, containing 1X PCR buffer, 4mM MgCl2, 10mM of each dNTP, 0.2 U GoTaq® G2 Flexi DNA polymerase, primers 10 µM and 1 µL of gDNA. The amplification of *rbcL* gene was performed under the following conditions: denaturation initial 94°Cx 2 min, 35 cycles, including an initial denaturation of 94°C x 30 seconds, alignment at 57°C x 30 seconds, extension at 72°C x 90 seconds, final extension at 72°C x 5 minutes. 5µL of each

PCR product was evaluated at agarose gel 2% with RedGel™ nucleotide staining, then, 15µL of PCR product was purified with E.Z.N.A. ® Cycle Pure Kit D6493-01 from OMEGA bio-tek.

Purified PCR products were sequenced by Macrogen (http://dna.macrogen.com/eng/). All sequences obtained were evaluated at Chromas (http://technelysium.com.au/wp/chromas) to clean chromatograms. Preliminary identification of the obtained sequences was done by performing a nucleotide BLAST search (blastn) (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

To examine genetic diversity and establish phylogenetic relation between *rbcL* gene sequences of *D. geminata*, a phylogenetic analysis was carried out including previously described diatom sequences for this gene in *D. geminata*. Thus, we examined all the 19 sequences described in this work, together with the seven sequences reported in Chile from Jaramillo et al. (2015) and 2 sequences available for *rbc*L gen in GenBank, reported by Nakov et al. (2014). Also, we examined sequences available at GenBankl for three additional genera of diatoms: *Gomphonema*, *Cymbella* and *Encyonema*. As a result, we examined a total of 48 sequences corresponding to 20 species of diatoms. These sequences were used to construct phylogenetic relationships trees using three approaches. The first one corresponded to a phenetic neighbour joining tree (NJ), which grouped sequences according to their degree of similarity or genetic distance (Saitou and Nei, 1987). The following two trees were built by using cladistics approximations, where group sequences were formed based on the degree of similarity to the previously identified groups. The cladistics trees were built using maximum parsimony (MP)

and maximum likelihood (ML) (REFERENCE). It should be noted that the latter is a probabilistic method, which is why it was selected for the final interpretation of phylogenetic relationships. To perform diversity analysis or genetic variability at the population level, the sequences were aligned using the ClustalW routine in the software package Molecular Evolutionary Genetic Analysis version 7.0 for bigger datasets (MEGA7) (Kumar, Stetcher and Kimura 2016). Once aligned, the sequences were collapsed into haplotypes using the FaBox package (Villesen, 2007) (http: // users-bir-c.au.dk/biopv/php/fabox/).

After the phylogenetic relationships identification of these species of Diatoms, sequences corresponding to samples obtained in the fluvial systems in Chile (7 reported in Jaramillo et al (2015) and 19 reported in this work) were analyzed to examine the population variability of *D. geminata* that cover the territory invasion of this specie in our country. To perform diversity analysis or genetic variability at the population level, the sequences were aligned using the ClustalW routine in the software package Molecular Evolutionary Genetic Analysis version 7.0 for bigger datasets (MEGA7) (Kumar, Stetcher and Kimura 2016). Once aligned, the sequences were collapsed into haplotypes using the FaBox package (Villesen, 2007) (http: // users-bir-c.au.dk/biopv/php/fabox/). This allowed us to enumerate the amount of haplotypes presents a level of the basins sampled, and also to determine the population genetic diversity, through the calculation of the index of diversity of Shannon-Wiener: 'H= - S piln(pi) considering the index i along the haplotypes identified in the previous section.

To infer the historical processes that gave rise to sample populations, the

relationships between haplotypes and geography were examined by a network of haplotypes.

This network was built using the R pegas 0.9 module (Paradis 2010), in R sofware (R Development Core Team, 2014). The haplotype network is a bi-directional graph that connects the different haplotypes identified, using a model of infinite DNA sequence sites (that is, uncorrected distance or Hamming), taking into account the deletion or omission of missing data pairs (Templeton et al., 1992). The size of the nodes in the resulting network is proportional to the number of sequences for each given haplotype, and the identity of the sequences present in each haplotype is represented by colors.

In addition, the number of estimated mutations between linked haplotypes is represented by small points. Finally, to exam the geographic variation in the identity and frequency haplotype, the geographical location of all sequences was represent by color coding the membership of the sequences in different groups of haplotypes.

Results

Phylogenetic relationships

Figures 1 to 3 illustrate the results obtained by performing a phylogenetic analysis of the 48 diatom sequences using different algorithms to construct the tree of phylogenetic relationships. In Figure 1 we report the results obtained by using a

phenetic algorithm, specifically, nearest neighbor linkage. In this result it can be seen that the genera *Gomphonema*, *Cymbella* and *Encyonema* are clearly separated from most of the sequences of the genus *Didymoshpenia*. Exceptions to this separation are *Didymosphenia* sequences corresponding to *Didymosphenia geminata* with Genebank accession code KJ011820, and KJ011818, reported in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) by Nakov et al. (2014), which both correspond to Lake Baikal, Siberia. Thus, the invasive *D. geminata* lineages observed in Chile do not correspond to the Siberian lineages described by Nakov et al. (2014).

This separation of the samples obtained in Chile with respect to the other sequences is also observed when examining the topologies obtained by the algorithms of maximum parsimony (cladistic analysis) and maximum likelihood (cladistic analysis) (Figures 2 and 3). Again, we also observed divergences between the siberian *D. geminata* sequences KJ011820 and KJ011818 with respect to those obtained in the present project and in the work of Jaramillo *et al*. (2015). Furthermore, despite the differences in phylogeny reported in our analysis, a consistent result is the observed separation of the samples collected in Chile with respect to the samples of *Gomphonema*, *Cymbella* and *Encyonema*.

Diversity and Population Structure

Regarding the population structure and its genetic variability, the results indicate that, independently of the method of analysis, a group of 15 to 18 sequences of

194 *Didymoshpenia geminata* are grouped with a very low genetic distance between
195 them. This suggests that there are a large number of the 26 Chilean samples that
196 share very similar sequences. Regardless of the algorithm selected, separations
197 can be seen in the remaining set of samples. These results indicate that there is an
198 important group of populations that could belong to a single lineage. In turn, the
199 remaining sequences could correspond to lineages that originated by mutation
200 from a first introduction or may correspond to new lineages that have been
201 introduced independently.
202
203 The analysis of haplotypes allowed to solve the high degree of similarity among the
204 26 sequences collected in Chile, particularly for those that do not appear as
205 differentiated in the phylogenetic trees (Figures 1 to 3). Table 1 details information
206 obtained for each sequence after using the FaBox software, including the result for
207 the identification of haplotypes based on *rbc*L gene from *Dydimosphenia geminata*
208 in Chile. The analysis of haplotypes showed that there are two frequent haplotypes
209 corresponding to haplotype I and IX, with 12 and 3 populations respectively, while
210 each of the remaining haplotypes presents a single population (Table 2). When
211 examining the genetic diversity for the different basins, there is a decrease along
212 the latitudinal gradient, with the Bío Bío basin being the most diverse, with six
213 haplotypes present, while the basins of the Puelo and Baker rivers being the least
214 diverse, presenting only a single haplotype (Figure 4). This could indicate a source
215 of invasion from the north to the south. In this sense, and considering only the
216 number of observed haplotypes, a first hypothesis formulated would be that the
217 observed genetic diversity could correspond to a single introduction of a diverse

218 population, with a point of origin in Bío Bío and later transfer or propagation
219 towards the South. However, there are alternative interpretations or hypotheses.
220 The first is that the high number of haplotypes could correspond to multiple
221 introductions, subsequent to the original introduction. Alternatively, a number of
222 these introductions could have given rise to new haplotypes, through mutations,
223 resulting in greater diversity, as observed in the Bío Bío, Valdivia, Bueno, Palena
224 and Aysén basins.
225
226 To examine to what degree the evidence obtained suggests one or another
227 hypothesis, the degree of similarity between haplotypes was assessed by a
228 network or graph analysis. Figure 5 depicts the haplotype network obtained from
229 the sequences (Paradis 2010). This network corresponds to a bi-directional graph
230 that connects the different haplotypes identified from the analyzed sequences. In
231 the network, each node corresponds to a population or sequence, and is
232 connected by a link to that (or those) haplotypes of greater similarity as to its
233 genetic sequence. The graph uses an uncorrected or Hamming distance,
234 corresponding to a model that assumes the existence of infinite DNA sequence
235 sites (Templeton et al., 1992). This network represents the number of sequences
236 for each haplotype given by the size of the node, also detecting by color the
237 identity of the sequences present in each haplotype. The points located along each
238 link represent the number of mutations between the haplotypes connected by said
239 link.

240 When examining the haplotype network, it can be seen that haplotype I, which is

241 the most frequent, is separated from haplotypes III, XI and XII by a single mutation.
242 Of these three, only haplotype III has very similar sequences, highlighting
243 haplotype VIII (one difference mutation with III), haplotype VII (two difference
244 mutations with III) and the IX population (two difference mutations with III).
245 Therefore, it could be expected that haplotype I corresponds to an introduction and
246 possibly IX also corresponds to an introduction. In order to identify these
247 possibilities more clearly, the pattern of geographic distribution of these haplotypes
248 was examined, given that space can significantly complement relationships. When
249 examining the phylogeographic variation of the genetic diversity of the *rbc*L gene of
250 *D. geminata* in Chile, we could see that the sequence I is widely distributed, while
251 the others have a bounded distribution, with the exception of the sequence IX
252 (Figure 6). However, the pattern of distance in number of mutations indicates that
253 at least haplotypes XI and XII correspond to mutations of haplotype I. On the other
254 hand, haplotypes II, IV, V, VI and XIII present great distances or accumulations of
255 mutations with respect to their haplotypes closer and as well with to haplotype I, at
256 least 4 of which are sequences of relatively isolated and little diverse populations.
257 All these antecedents suggest a potential for at least two events of invasion of the
258 plague *D. geminata* in Chile, being able to reach up to ten events of introduction
259 from different lineages.

260

261 **Discussion and Conclusion**

262 Regarding the genetic variability of *Didymosphenia geminata* samples reported in

this work, we highlight three points that merit discussion. The first corresponds to the quality and quantity of the results obtained, and their relevance in terms of population screening as well as phylogenetic or estimation of relations between diatom lineages. The second point corresponds to the genetic variability observed, and the degree of resolution at the population and phylogenetic level, evidenced mainly in the phylogenetic trees generated. Finally, the third point relates to the implications of inferred haplotypes from the sequences studied, particularly for the invasion and potential sources of this species in Chile.

With regard to the first point, it should be noted that the process of extraction and purification of DNA for this group of organisms is a complex process and is currently being actively researched worldwide. In this regard, our results corroborate the validity and reliability of the *rbc*L gene as a suitable molecular marker to extract useful genetic information for the analysis of this specie. Similarly, it should be noted that the results obtained contribute increasing the amount of evidence available for this gene in the genus *Didymosphenia* worldwide, doubling the number of sequences reported of this gene. Both elements provide an important degree of validity and reliability in the molecular protocol performed. In this context, the variation introduced by our results on the estimation of the total diatoms phylogeny, is a point of particular interest in the study of this group in general. The great extent of the specie *D. geminata* and its high invasive potential require a comprehensive approach that considers in particular molecular markers to elucidate kinship relationships, especially if it is desired to identify potential sources of the different invading lineages, both in Chile and in other countries. On

the other hand, the validity and reliability of molecular markers such as *rbc*L will allow broaden the range of tools available for the taxonomy of this pest, helping to support existing identifications based on the taxonomy from morphological features.

A second important result is the clear difference between the samples collected in Chile, with respect to the sequences available from other species. In this sense, the results strongly suggest that the samples of *D. geminata* reported by Nakov *et al*. (2014) would not correspond to the population source of the current invasion of this specie in Chile. This makes relevant and necessary the development of additional sampling and sequencing efforts, both in our country and in potential sources, like in South America and in other continents. In addition, to successfully solving the genetic differences of *D. geminata* strains present in Chile, the three phylogenetic analysis methods used in this work coincide in indicating that a large number of the populations sequenced in our country correspond to sequences that are not very variable. Less than half of the remaining samples can be solved, presenting differences according to the agglomeration methods used. This indicates that there is an important degree of population diversification in this diatom specie.

Finally, we discuss the implications of this genetic variability for the interpretation and management of the invasion process of this pest. The presence of a high number of haplotypes, many of them very different, suggests that a number greater than 1 or 2 invasions could be feasible. However, studies with more molecular evidence are required, expanding the spectrum of genetic markers and the

309 populations to include, hopefully encompassing potentially population sources. In
310 any case, the results indicate that the genetic structure of this specie is dominated
311 by a dominant haplotype, defined as I, found in most of the basins studied. In turn,
312 the large number of existing base substitutions for most of the remaining
313 haplotypes suggests that a high number of lineages have invaded Chilean basins.
314 In particular, at least 4 or 5 of these lineages are found in basins without other
315 haplotypes. This evidence may be consistent several alternative hypotheses. First,
316 the ongoing invasion could be explained by multiple invasion events, which would
317 require long-distance dispersal events from diverse sources. Likely scenarios could
318 include either transport from existing invaded sites in South America, or transport
319 from other sites in North America. Alternatively, initial invasions could have
320 contained a mixed set of genetic lineages, followed by long-distance dispersal
321 events. In this regard, the spatial location of the different haplotypes strongly
322 supports the existence of short- and large-scale transport mechanisms, given the
323 geographical extension of the two most frequent haplotypes, and the large
324 geographic dispersion of the rarer haplotypes.

325

326 On the other hand, an alternative mechanism that could account for the observed
327 pattern of haplotype diversification could be the mutation of initial invasive lineages
328 after they were introduced to Chilean river basins. This would require a large
329 number of generations and a short generation time. A comprehensive review of
330 available literature indicated that information on the generation times of freshwater
331 diatoms is sparse (Kilroy 2004). Growth rates measured under laboratory culture

conditions have been found to vary considerably among species, as well as with temperature and light intensity, with generation times being quite rapid (< 30 h at temperatures between 12 and 20°C) in many cases (Baars, 1983). The large freshwater species *Pinnularia gibba* presented the shortest generation time recorded, with cells dividing every 13 h under the highest light – temperature (20°C) combination (Baars 1983). While there's no specific information on the generation times of *Didymosphenia*, cell division every 30 h or less could lead to rapid accumulation of biomass, as well as eventually allowing rapid genetic change to accumulate. Again, elucidating these alternative hypotheses would require the analysis of a larger number of samples in the national territory, together with samples from other countries that may be potential sources for the Chilean invasion of *D. geminata*.

352 References

353

354 AMAKAIK Consultoría Ambiental (2016). Manual para el monitoreo e
355 identificación de la microalga bentónica *Didymosphenia geminata*. 2da
356 Edición. Subsecretaría de Pesca, Gobierno de Chile.

357

358 Bothwell M.L, Lynch D.R, Wright H. & Deniseger J. (2009) On the Boots of
359 Fishermen: The History of Didymo Blooms on Vancouver Island, British
360 Columbia, *Fisheries*, 34:8, 382-388,

361

362 Carolina A. Díaz P., Ximena Molina P. & Vivian Montecino B (2012). Manual
363 para el Monitoreo e Identificación de la microalga bentónica *Didymosphenia*
364 *geminata*. Subsecretaría de Pesca, Gobierno de Chile.

365

366 Jaramillo A, Osman D, Caputo L, Cárdenas L (2015) Molecular evidence of
367 a *Didymosphenia geminata* (Bacillariophyceae) invasion in Chilean
368 freshwater systems. *Harmful Algae* 49: 117-123.

369

370 Kilroy C. (2004). A new alien diatom, Didymosphenia geminata (Lyngbye)
371 Schmidt: its biology, distribution, effects and potential risks for New Zealand
372 fresh waters. Report from National Institute of Water & Atmospheric
373 Research Ltd. NIWA Client Report: CHC2004-128. November 2004. NIWA
374 Project: ENS05501.

Kumar S, Stecher G, and Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874

Leone, P.B., Cerda, J., Sala, S., Reid, B., 2014. Mink (Neovison vison) as a natural vector in the dispersal of the diatom *Didymosphenia geminata*. *Diatom Res*. 29 (3), 259–266.

Nakov,T., Ruck,E.C., Galachyants,Y., Spaulding,S.A. & Theriot,E.C. (2014) Molecular phylogeny of the *Cymbellales (Bacillariophyceae*, *Heterokontophyta*) with a comparison of models for accommodating rate variation across sites. *Phycologia* 53 (4), 359-373.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419-420.

R Development Core Team. 2014. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Tamura K., Nei M., and Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)* 101:11030-11035.

399 Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for
400 reconstructing phylogenetic trees. Molecular biology and evolution, 4(4),
401 406-425.
402
403 Segura, P., 2011. A slimy invader blooms in the rivers of Patagonia. Science
404 331, 17–18.
405
406 Templeton, A. R., Crandall, K. A. and Sing, C. F. (1992) A cladistics analysis
407 of phenotypic association with haplotypes inferred from restriction
408 endonuclease mapping and DNA sequence data. III. *Cladogram estimation.*
409 *Genetics*, 132, 619–635.
410
411 Villesen P. 2007. FaBox: an online toolbox for fasta sequences. *Molecular*
412 *Ecology Notes* 7: 965–968.
413
414
415

Figure 1

Phylogenetic relationship of *Dydimosphenia geminata* and other related diatom species based on the *rbc*L gene.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27330v1 | CC BY 4.0 Open Access | rec: 8 Nov 2018, publ: 8 Nov 2018

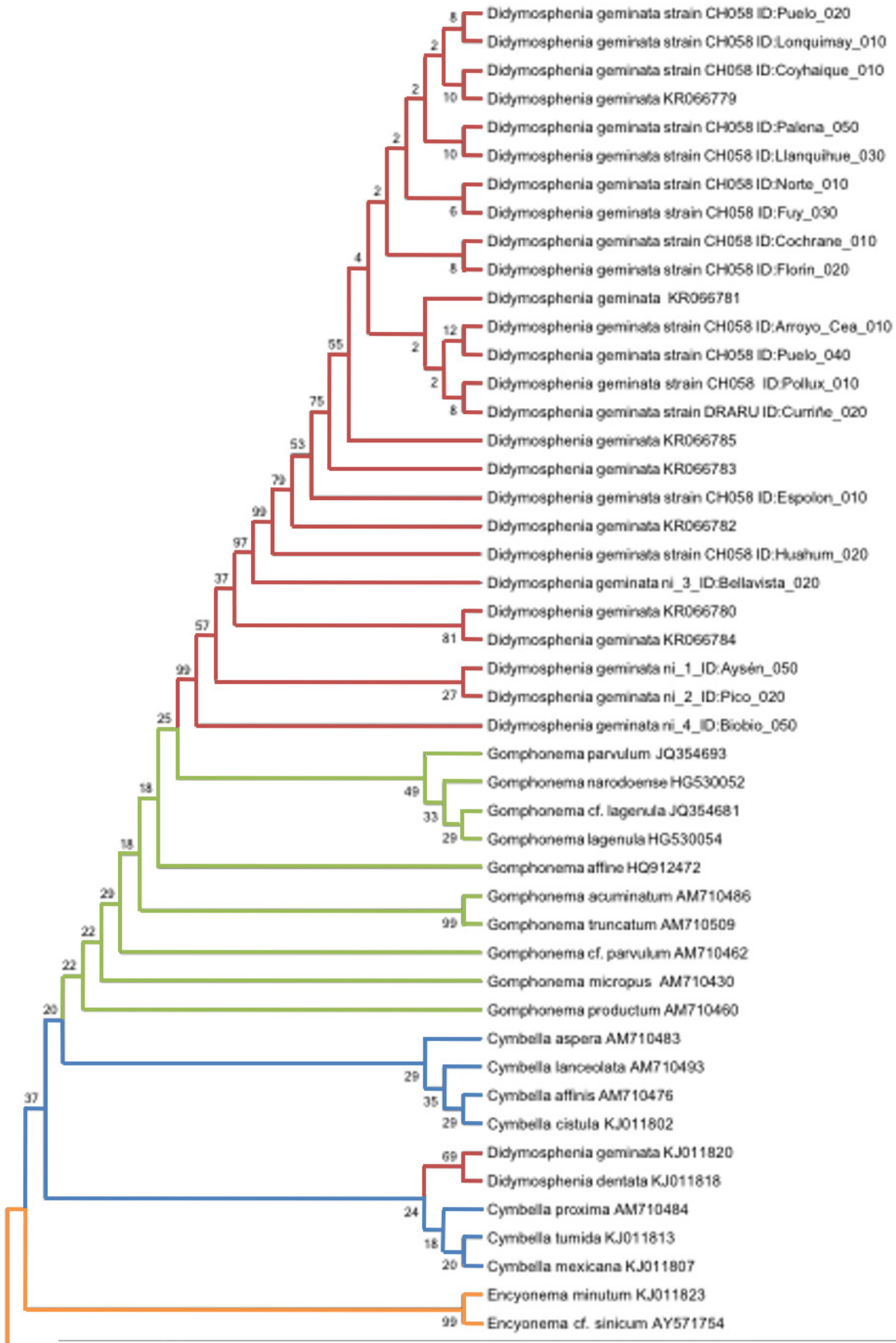

Didymosphenia geminata strain CH058 ID:Puelo_020
Didymosphenia geminata strain CH058 ID:Lonquimay_010
Didymosphenia geminata strain CH058 ID:Coyhaique_010
Didymosphenia geminata KR066779
Didymosphenia geminata strain CH058 ID:Palena_050
Didymosphenia geminata strain CH058 ID:Llanquihue_030
Didymosphenia geminata strain CH058 ID:Norte_010
Didymosphenia geminata strain CH058 ID:Fuy_030
Didymosphenia geminata strain CH058 ID:Cochrane_010
Didymosphenia geminata strain CH058 ID:Florin_020
Didymosphenia geminata KR066781
Didymosphenia geminata strain CH058 ID:Arroyo_Cea_010
Didymosphenia geminata strain CH058 ID:Puelo_040
Didymosphenia geminata strain CH058 ID:Pollux_010
Didymosphenia geminata strain DRARU ID:Curriñe_020
Didymosphenia geminata KR066785
Didymosphenia geminata KR066783
Didymosphenia geminata strain CH058 ID:Espolon_010
Didymosphenia geminata KR066782
Didymosphenia geminata strain CH058 ID:Hualhum_020
Didymosphenia geminata ni_3_ID:Bellavista_020
Didymosphenia geminata KR066780
Didymosphenia geminata KR066784
Didymosphenia geminata ni_1_ID:Aysén_050
Didymosphenia geminata ni_2_ID:Pico_020
Didymosphenia geminata ni_4_ID:Biobio_050
Gomphonema parvulum JQ354693
Gomphonema narodoense HG530052
Gomphonema cf. lagenula JQ354681
Gomphonema lagenula HG530054
Gomphonema affine HQ912472
Gomphonema acuminatum AM710486
Gomphonema truncatum AM710509
Gomphonema cf. parvulum AM710462
Gomphonema micropus AM710430
Gomphonema productum AM710460
Cymbella aspera AM710483
Cymbella lanceolata AM710493
Cymbella affinis AM710476
Cymbella cistula KJ011802
Didymosphenia geminata KJ011820
Didymosphenia dentata KJ011818
Cymbella proxima AM710484
Cymbella tumida KJ011813
Cymbella mexicana KJ011807
Encyonema minutum KJ011823
Encyonema cf. sinicum AY571754

Figure 2

Phylogenetic relationship of *Dydimosphenia geminata* and other related species of diatoms based on the *rbcL* gene.

The results obtained on the basis of cladistics algorithms of maximum parsimony are illustrated. For each branch, the percentages of bootstrap obtained are indicated. The value to the right of each branch indicates the later probabilities. The represented diatom genera are identified by the colors of the corresponding branch. In orange trees the species of the genus *Encyonema* are highlighted, whereas the species of *Gomphonema* are shown in green. Blue shows those sequences that correspond to species of *Cymbella*, and in red those sequences that correspond to species of the genus *Didymosphenia*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27330v1 | CC BY 4.0 Open Access | rec: 8 Nov 2018, publ: 8 Nov 2018

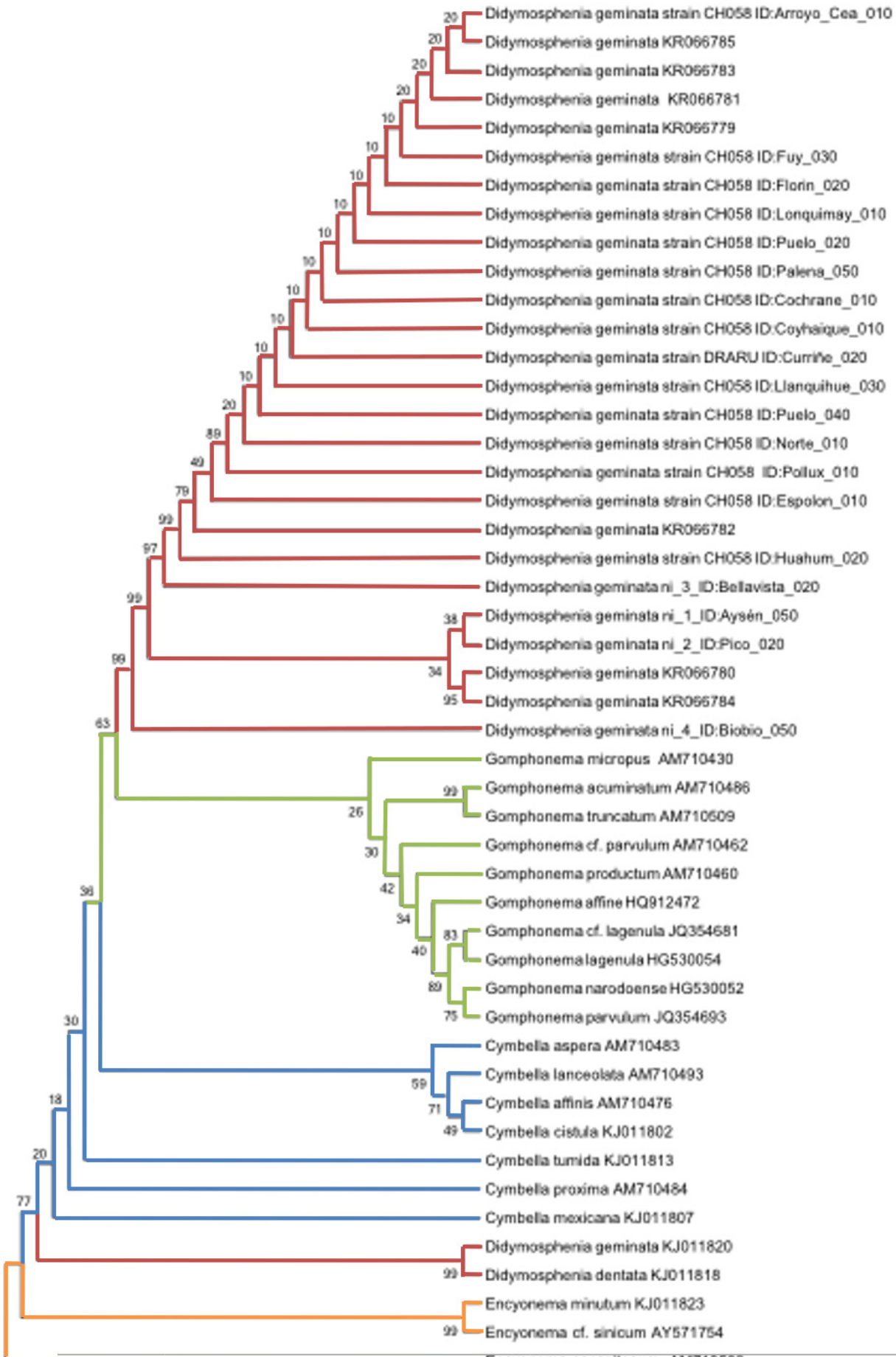

Didymosphenia geminata strain CH058 ID:Arroyo_Cea_010
Didymosphenia geminata KR066785
Didymosphenia geminata KR066783
Didymosphenia geminata KR066781
Didymosphenia geminata KR066779
Didymosphenia geminata strain CH058 ID:Fuy_030
Didymosphenia geminata strain CH058 ID:Florin_020
Didymosphenia geminata strain CH058 ID:Lonquimay_010
Didymosphenia geminata strain CH058 ID:Puelo_020
Didymosphenia geminata strain CH058 ID:Palena_050
Didymosphenia geminata strain CH058 ID:Cochrane_010
Didymosphenia geminata strain CH058 ID:Coyhaique_010
Didymosphenia geminata strain DRARU ID:Curiñe_020
Didymosphenia geminata strain CH058 ID:Llanquihue_030
Didymosphenia geminata strain CH058 ID:Puelo_040
Didymosphenia geminata strain CH058 ID:Norte_010
Didymosphenia geminata strain CH058 ID:Pollux_010
Didymosphenia geminata strain CH058 ID:Espolon_010
Didymosphenia geminata KR066782
Didymosphenia geminata strain CH058 ID:Huahum_020
Didymosphenia geminata ni_3_ID:Bellavista_020
Didymosphenia geminata ni_1_ID:Aysén_050
Didymosphenia geminata ni_2_ID:Pico_020
Didymosphenia geminata KR066780
Didymosphenia geminata KR066784
Didymosphenia geminata ni_4_ID:Biobio_050
Gomphonema micropus AM710430
Gomphonema acuminatum AM710486
Gomphonema truncatum AM710609
Gomphonema cf. parvulum AM710462
Gomphonema productum AM710460
Gomphonema affine HQ912472
Gomphonema cf. lagenula JQ354681
Gomphonema lagenula HG530054
Gomphonema narodoense HG530052
Gomphonema parvulum JQ354693
Cymbella aspera AM710483
Cymbella lanceolata AM710493
Cymbella affinis AM710476
Cymbella cistula KJ011802
Cymbella tumida KJ011813
Cymbella proxima AM710484
Cymbella mexicana KJ011807
Didymosphenia geminata KJ011820
Didymosphenia dentata KJ011818
Encyonema minutum KJ011823
Encyonema cf. sinicum AY571754

Figure 3

Phylogenetic relationship of *Dydimosphenia geminata* and other related species of diatoms based on the *rbcL* gene.

The results obtained are based on the probabilistic cladistics algorithm of maximum likelihood. The percentages of bootstrap obtained are indicated for each branch. The value to the right of each branch indicates the later probabilities. The represented diatom genera are identified by the colors of the corresponding branch. In orange trees the species of the genus *Encyonema* are highlighted, whereas the species of *Gomphonema* are shown in green. Blue shows those sequences that correspond to species of *Cymbella*, and in red those sequences that correspond to species of the genus *Didymosphenia*.

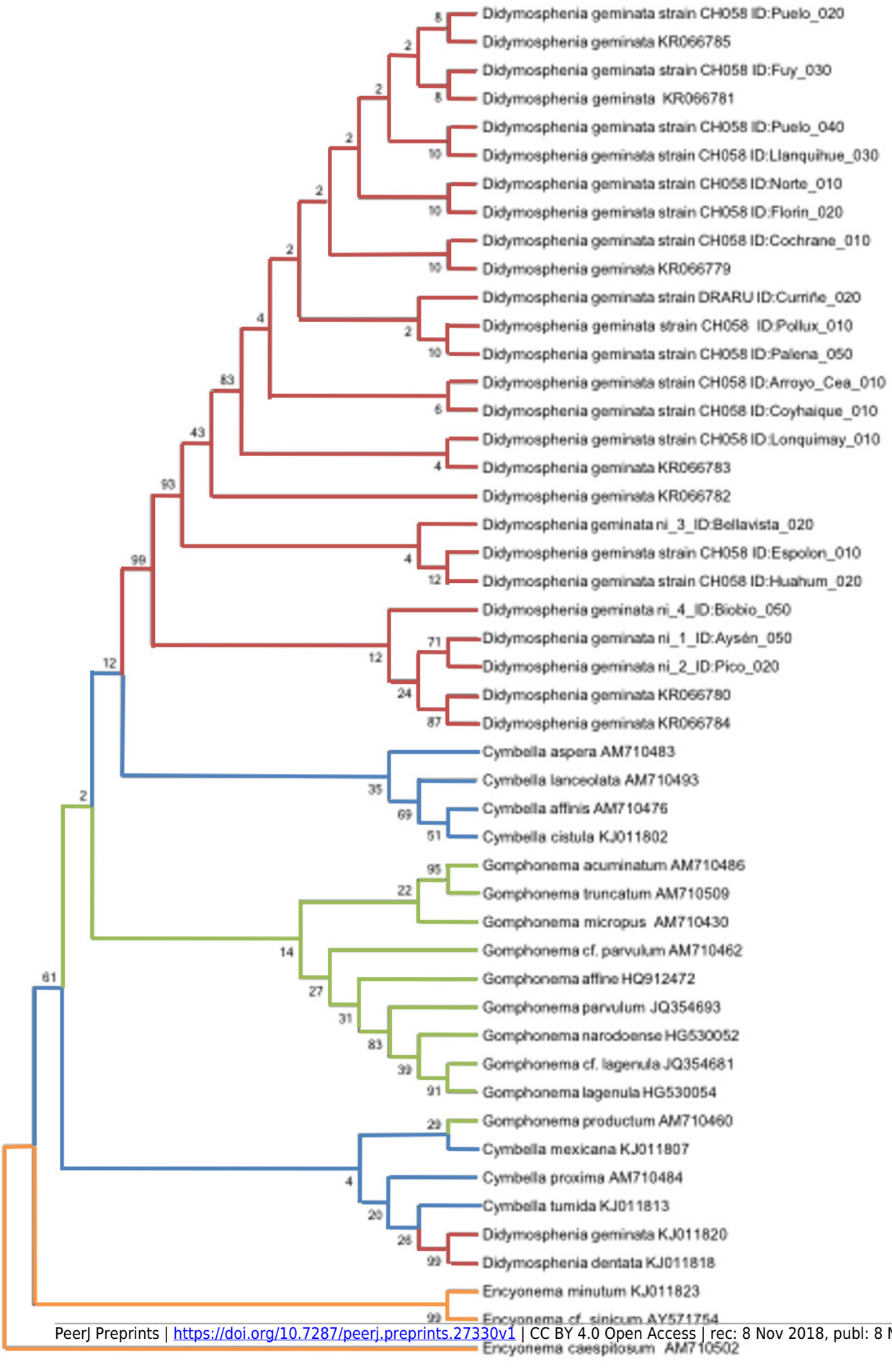

Didymosphenia geminata strain CH058 ID:Puelo_020
Didymosphenia geminata KR066785
Didymosphenia geminata strain CH058 ID:Fuy_030
Didymosphenia geminata KR066781
Didymosphenia geminata strain CH058 ID:Puelo_040
Didymosphenia geminata strain CH058 ID:Llanquihue_030
Didymosphenia geminata strain CH058 ID:Norte_010
Didymosphenia geminata strain CH058 ID:Florin_020
Didymosphenia geminata strain CH058 ID:Cochrane_010
Didymosphenia geminata KR066779
Didymosphenia geminata strain DRARU ID:Curriñe_020
Didymosphenia geminata strain CH058 ID:Pollux_010
Didymosphenia geminata strain CH058 ID:Palena_050
Didymosphenia geminata strain CH058 ID:Arroyo_Cea_010
Didymosphenia geminata strain CH058 ID:Coyhaique_010
Didymosphenia geminata strain CH058 ID:Lonquimay_010
Didymosphenia geminata KR066783
Didymosphenia geminata KR066782
Didymosphenia geminata ni_3_ID:Bellavista_020
Didymosphenia geminata strain CH058 ID:Espolon_010
Didymosphenia geminata strain CH058 ID:Huahum_020
Didymosphenia geminata ni_4_ID:Biobio_050
Didymosphenia geminata ni_1_ID:Aysén_050
Didymosphenia geminata ni_2_ID:Pico_020
Didymosphenia geminata KR066780
Didymosphenia geminata KR066784
Cymbella aspera AM710483
Cymbella lanceolata AM710493
Cymbella affinis AM710476
Cymbella cistula KJ011802
Gomphonema acuminatum AM710486
Gomphonema truncatum AM710509
Gomphonema micropus AM710430
Gomphonema cf. parvulum AM710462
Gomphonema affine HQ912472
Gomphonema parvulum JQ354693
Gomphonema narodoense HG530052
Gomphonema cf. lagenula JQ354681
Gomphonema lagenula HG530054
Gomphonema productum AM710460
Cymbella mexicana KJ011807
Cymbella proxima AM710484
Cymbella tumida KJ011813
Didymosphenia geminata KJ011820
Didymosphenia dentata KJ011818
Encyonema minutum KJ011823
Encyonema cf. sinicum AY571754
Encyonema caespitosum AM710502

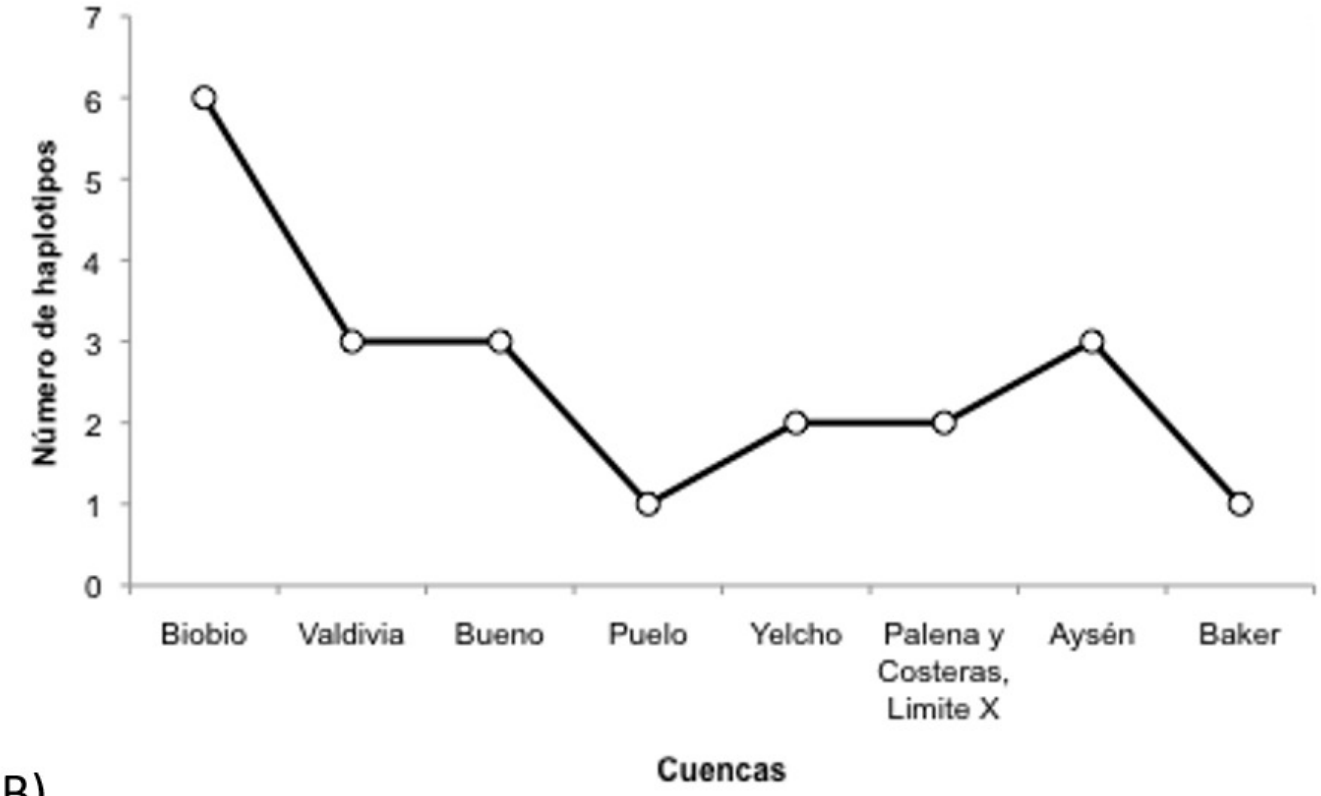
Figure 4

Population genetic diversity of Dydimosphenia geminata in Chile.

Variation at the basin level for the diversity of haplotypes of the rbcL gene of *Dydimosphenia geminata* in Chile is illustrated. The number of haplotypes (A) and Shannon diversity (B) are illustrated for the 8 basins studied.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27330v1 | CC BY 4.0 Open Access | rec: 8 Nov 2018, publ: 8 Nov 2018

(A)


Número de haplotipos
0
1
2
3
4
5
6
7
Biobio
Valdivia
Bueno
Puelo
Yelcho
Palena y Costeras, Limite X
Aysén
Baker
Cuencas

(B)

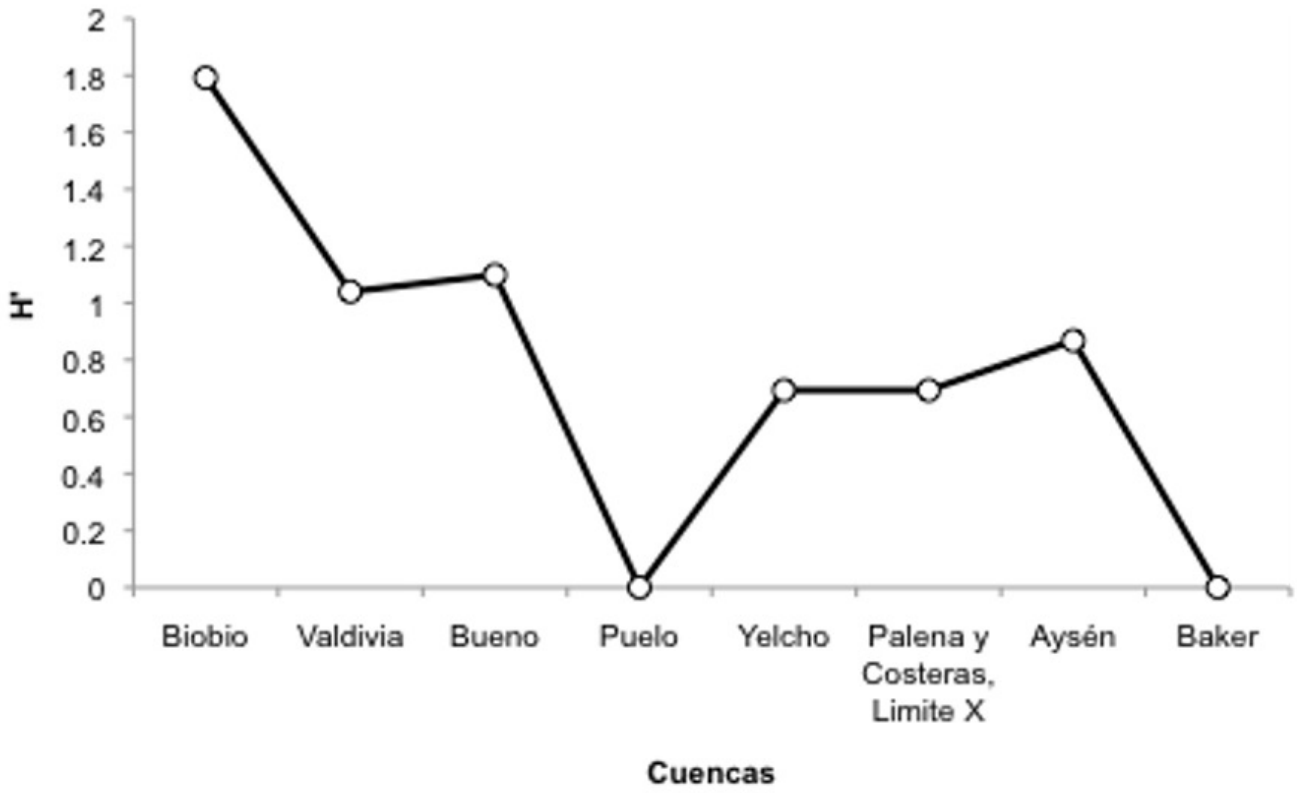

H'
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Biobio
Valdivia
Bueno
Puelo
Yelcho
Palena y Costeras, Limite X
Aysén
Baker
Cuencas

Figure 5

Median binding network of the haplotypes of the *rbcL* gene of *Dydimosphenia geminata* in Chile.

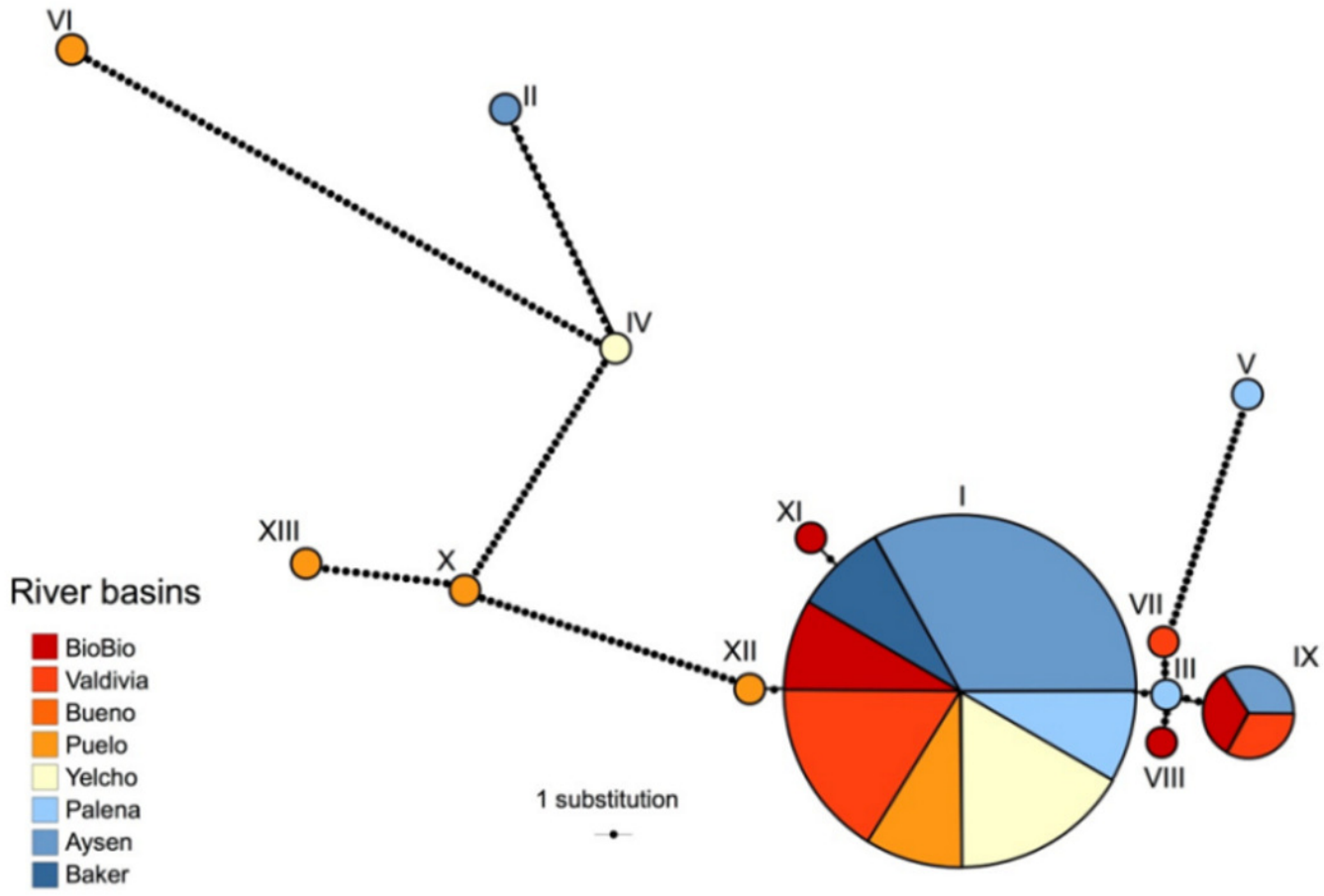
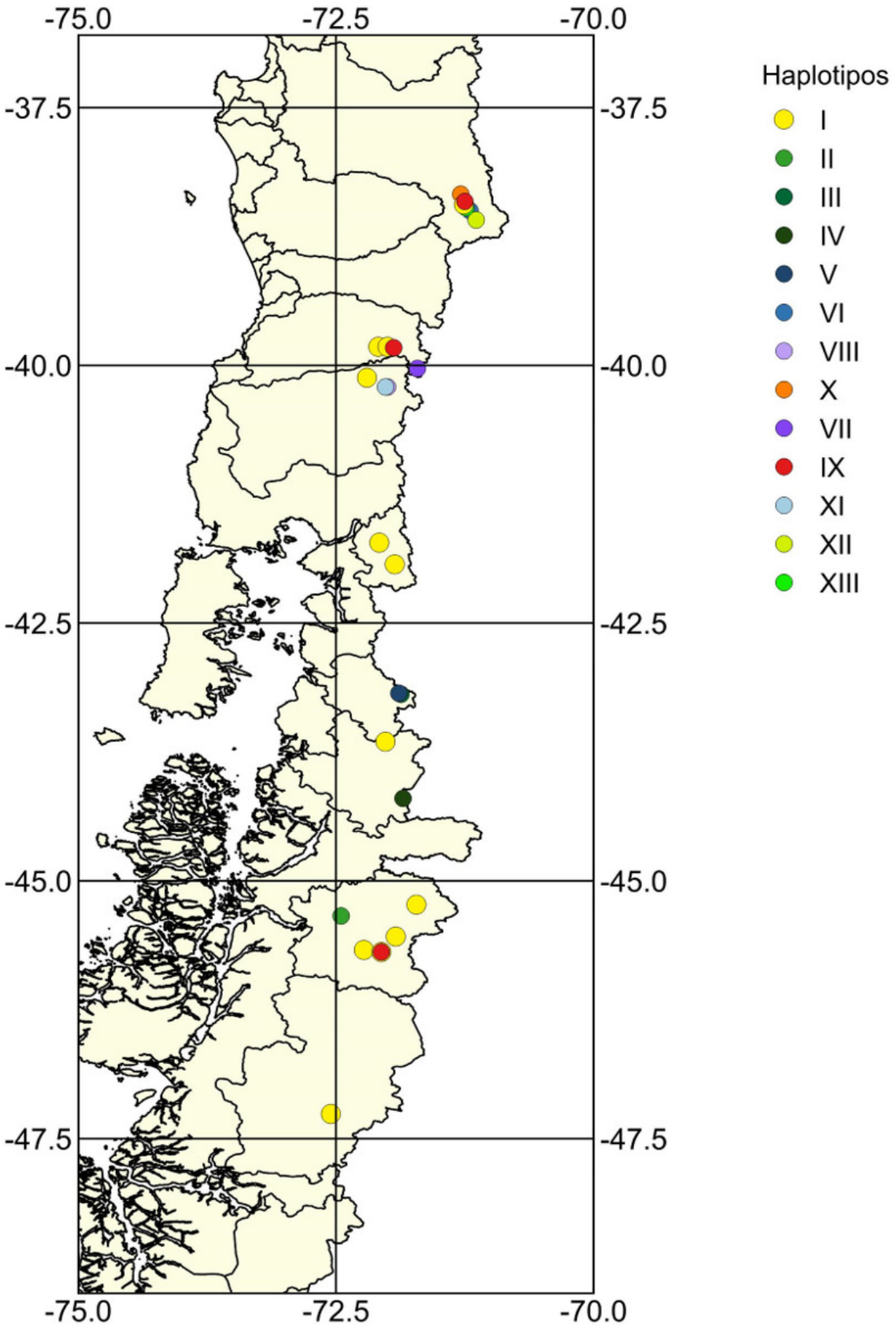


Figure 6

Phylogeographic variation of the genetic diversity of the *rbcL* gene of *Dydimosphenia geminata* in Chile.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27330v1 | CC BY 4.0 Open Access | rec: 8 Nov 2018, publ: 8 Nov 2018

Table 1(on next page)

Sampling sites, where samples were taken, including region, basin, river and geographic coordinates.

Each sample was designated with a Final Name for identification.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27330v1 | CC BY 4.0 Open Access | rec: 8 Nov 2018, publ: 8 Nov 2018

1

Region	Basin	River	Final Name	Latitude	Longitude
Araucanía	R.Biobío	R.Lonquimay	Lonquimay_010	38º 26' 31,349" S	71º 15' 40,3" W
Araucanía	R.Biobío	R.Biobío	Biobío_050	38º 30' 12,141" S	71º 12' 11,803" W
Los_Ríos	R.Valdivia	R.Llanquihue	Llanquihue_030	39º 49' 21,318" S	72º 5' 12,217" W
Los_Ríos	R.Valdivia	R.Fuy	Fuy_030	39º 49' 27,674" S	71º 59' 48" W
Los_Ríos	R.Valdivia	R.Huahum	Huahum_020	40º 2' 0,401" S	71º 42' 49,713" W
Los_Ríos	R.Bueno	R.Caunahue	Florín_020	40º 7' 15,662" S	72º 12' 6,251" W
Los_Ríos	R.Bueno	R.Currine	Currine_020	40º 12' 49,913" S	72º 0' 8,715" W
Los_Lagos	R.Puelo	R.Puelo	Puelo_040	41º 43' 5,95" S	72º 4' 44,418" W
Los_Lagos	R.Puelo	R.Puelo	Puelo_020	41º 56' 4,477" S	71º 55' 36,428" W
Los_Lagos	R.Yelcho	R.Espolón	Bellavista_020	43º 10' 44,384" S	71º 53' 15,687" W
Los_Lagos	R.Yelcho	R.Espolón	Espolón_010	43º 11' 35,324" S	71º 52' 14,516" W
Los_Lagos	R.Palena_y_Costeras_Limite_Décima_Región	R.Palena	Palena_050	43º 38' 57,546" S	72º 1' 17,528" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Palena_y_Costeras_Limite_Décima_Región	R.Pico	Pico_020	44º 11' 55,105" S	71º 51' 1,25" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Aysén	R.Norte	Norte_010	45º 14' 2,294" S	71º 43' 19,44" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Aysén	R.Coyhaique	Aysén_050	45º 20' 41,333" S	72º 27' 13,656" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Aysén	R.Coyhaique	Coyhaique_010	45º 32' 24,866" S	71º 55' 3,102" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Aysén	R.Simpson	Arroyo_Cea_010	45º 40' 14,788" S	72º 13' 40,66" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Aysén	R.Pollux	Pollux_010	45º 41' 30,309" S	72º 3' 25,297" W
Aysén_del_Gral._Carlos_Ibanez_del_Campo	R.Baker	R.Cochrane	Cochrane_010	47º 15' 19,08" S	72º 32' 53,49" W

2

3

Table 2(on next page)

Location of the genetic sequence samples used to characterize the genetic diversity of *Didymosphenia geminata* in Chile.

<!--[if !supportAnnotations]--> <!--[endif]--> The names of the basins and rivers and the geographical location of each sampled population are detailed. In the corresponding case, access masters are indicated for the Genbank database (*rbc*L gene) for each sample. The code of each population identifies the samples represented in Figures 1 to 4. Samples are sorted according to the results of haplotype analysis performed using the Pegas module (Paradis, 2010), indicating the haplotype corresponding to each one of them.

Basin	River	Final Name	Lat (ºS)	Long (ºO)	Haplotype
Biobío	Lonquimay	Lonquimay_010	-38.44	-71.26	
Valdivia	Llanquihue	Llanquihue_030	-39.82	-72.09	
Valdivia	Fuy	Fuy_030	-39.82	-72.00	
Bueno	Florín	Florín_020	-40.12	-72.20	
Puelo	Puelo	Puelo_040	-41.72	-72.08	
Puelo	Puelo	Puelo_020	-41.93	-71.93	
Palena y C. Costeras, Límite X Región	Palena	Palena_050	-43.65	-72.02	I
Aysén	Norte	Norte_010	-45.23	-71.72	
Aysén	Coyhaique	Coyhaique_010	-45.54	-71.92	
Aysén	Cea	Arroyo_Cea_010	-45.67	-72.23	
Aysén	Pollux	Pollux_010	-45.69	-72.06	
Baker	Cochrane	Cochrane_010	-47.26	-72.55	
Aysén	Aysén	Aysén_050	-45.34	-72.45	II
Yelcho	Espolón	Espolón_010	-43.19	-71.87	III
Palena y C. Costeras, Límite X Región	Pico	Pico_020	-44.20	-71.85	IV
Biobío	Rahue	KR066785*	-38.41	-71.25	
Valdivia	Fuy	KR066779*	-39.83	-71.94	IX
Aysén	Pollux	KR066781*	-45.69	-72.06	
Yelcho	Espolón	Bellavista_020	-43.18	-71.89	V
Biobío	Biobío	Biobío_050	-38.50	-71.20	VI
Valdivia	Huahum	Huahum_020	-40.03	-71.71	VII
Bueno	Curriñe	Curriñe_020	-40.21	-72.00	VIII
Biobío	Caracoles	KR066780*	-38.34	-71.29	X
Bueno	Curriñe	KR066782*	-40.21	-72.02	XI
Biobío	Tucapel	KR066783*	-38.59	-71.14	XII
Biobío	Lollen	KR066784*	-38.47	-71.24	XIII

*Sequences reported by Jaramillo *et al.* (2015)