Brine chemistry matters: isolation by environment and by distance explain population genetic structure of Artemia franciscana in saline lakes

Dagmar Frisch, Christophe D Lejeusne, Masaki Hayashi, Mark T Bidwell, Javier Sánchez-fontenla, Andy J Green

Brine chemistry matters: isolation by environment and by distance explain population genetic structure of *Artemia franciscana* in saline lakes

Dagmar Frisch[1,*]

Christophe Lejeusne[2,3]

Masaki Hayashi[4]

Mark T. Bidwell[5]

Javier Sánchez-Fontenla[3]

Andy J. Green[3]

[1]University of Birmingham, School of Biosciences, Birmingham UK

[2]Aix Marseille Univ, CNRS, IRD, Avignon Université, IMBE, UMR 7263, Station Marine d'Endoume, Rue de la Batterie des Lions, 13007 Marseille, France

[3]Department of Wetland Ecology, Estación Biológica de Doñana (EBD-CSIC), C/Américo Vespucio 26, 41092 Sevilla, Spain

[4]Department of Geoscience, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada

[5]Canadian Wildlife Service, Environment and Climate Change Canada, Saskatoon, SK, Canada

* corresponding author, email: dfrisch@ou.edu

Running title: Population genetics of native *Artemia*

25 Keywords: brine shrimps, native region, hydrogeochemistry, colonization route, dispersal

Abstract

1. The American brine shrimp *Artemia franciscana* is important in aquaculture and has become invasive in other continents, aided by dispersal via waterbirds. However, little is known about processes underlying its genetic diversity and population structure in its natural habitat in North America. These processes, including dispersal and local adaptation, are pivotal drivers of species distribution and community structure, and therefore central to aquatic biodiversity.

2. We studied 15 populations in natural saline lakes of Saskatchewan, Canada to determine the influence of variation in geological history, water chemistry, lake size and location. We aimed to determine the relative importance of isolation by distance (IBD) and isolation by environment (IBE) using the cytochrome *c* oxidase subunit 1 gene (CO1) as a mitochondrial marker and five nuclear microsatellite markers.

3. Geographic patterns for CO1 and microsatellites differed, with lakes clustering in different groups based on genetic distances according to the marker used. CO1 better indicated historical colonization processes, suggesting potential routes of initial colonization when lakes were formed after deglaciation 11,000-15,000 years ago.

4. Differentiation between lakes based on nuclear markers was strongly related to variation in hydrochemistry, suggested by distance-based redundancy analysis (dbRDA), but there was no indication of IBD. The ratio between alkalinity and the sum of Ca and Mg concentrations was particularly important, although a lake with a high Cl concentration caused by Potash mining also had a unique *Artemia* population.

5. Geochemistry is important in the adaptive radiation of anostracan crustaceans. Our study suggests it also underlies intraspecific genetic variation between populations, promoting IBE, and making dispersal ineffective when cysts are moved by birds between lakes with different hydrochemistry.

Introduction

The role of dispersal and environmental filtering in structuring communities and populations is central to research that seeks a better understanding of species distributions and biogeographical patterns (D'Amen, Mod, Gotelli, & Guisan, 2018; Ovaskainen et al., 2017). Dispersal connects populations across spatial distances, leading to gene flow if followed by successful establishment in the new habitat (Bohonak, 1999). Such successful establishment may be hindered by unsuitable local environmental conditions or by a resident population with a numerical or fitness advantage resulting from priority effects, or local adaptation (De Meester et al 2016). Several patterns of genetic differentiation are predicted to result from three main processes or their interaction: isolation by distance (IBD (Wright, 1943)), isolation by environment (IBE (Wang & Bradburd, 2014) and isolation by adaptation (IBA (Nosil, Egan, & Funk, 2008)). The processes underlying these spatial genetic patterns can be identified by testing relationships between genetic differentiation and geographic and ecological distances (Orsini, Vanoverbeke, Swillen, Mergeay, & De Meester, 2013).

Here, we examine drivers of biogeographic patterns of the brine shrimp *Artemia franciscana* Kellog, 1906 (Branchiopoda: Anostraca) in its native range of North America. This species is highly invasive in other continents, to which it was originally exported from the Great Salt Lake and San Francisco Bay as resistant eggs (cysts) used widely as food in aquaculture (Amat et al., 2005; Horváth et al., 2018). It has become a model species in toxicology (Ruebhart et al., 2008; Libralato et al., 2016). In addition, it is also probably the best known example of an aquatic invertebrate that disperses via migratory waterbirds, which consume *Artemia* and disperse their cysts via gut passage owing to their ability to resist digestion (Muñoz, Amat, Green, Figuerola, & Gómez, 2013; Reynolds, Miranda, & Cumming, 2015; Sánchez, Hortas, Figuerola, & Green, 2012).

We focus our study on saline lakes of the Saskatchewan prairies, within the Great Plains bioregion for anostracans (Rogers, 2014a). This area was covered by an ice sheet during the last glacial period until the onset of deglaciation of southern Saskatchewan about 17,000 years BP, which gradually retreated in a north-east direction (Christiansen, 1979).

The export of cysts from San Francisco Bay and the Great Salt Lake in the USA has largely determined the genetic structure of *A. franciscana* in Mexico, as well as across continents other than North America (Eimanifar, Van Stappen, Marden, & Wink, 2014; Horváth et al., 2018). However, populations in central Canada, where the lakes freeze in winter and the hydrochemistry is markedly different, are unaffected by this export (Muñoz et al., 2013). Population genetic studies of this and other *Artemia* species have previously been concentrated in solar salt works or aquaculture ponds (Muñoz et al., 2008; Muñoz et al., 2013, 2014), and studies in natural lakes are rare. Naihong et al. (2000) found greater allozyme heterozygosity for *A. sinica* in larger lakes. Panmixia has been reported for *A. franciscana* in Great Salt Lake (Eimanifar, Marden, Braun, & Wink, 2015) and for *A. urmiana* in Lake Urmia (Eimanifar & Wink, 2013).

Within North America, the distribution of different species of anostracan crustaceans is strongly related to regional differences in geochemistry, and hence hydrochemistry (Rogers, 2014a). This suggests that variation in geo- and hydrochemistry may also be a strong selective force for local adaptation amongst widely distributed species such as *A. franciscana*, and may potentially result in population genetic patterns that suggest IBE. This is supported by differences observed under laboratory conditions between *A. franciscana* populations in their tolerance of different ionic concentrations (Bowen, Buoncristiani, & Carl, 1988). Saskatchewan has an estimated 500 saline lakes greater than one square kilometer in area, mainly dominated by sodium sulfate and magnesium

sulfate salts (Hammer, 1993). Beginning in 1918 in Muskiki Lake, sodium sulfate has been industrially extracted from 20 lakes, which supplied about 50% of the North American demand (Last & Ginn, 2005). However, the hydrochemistry varies considerably between different saline lakes in this region (Bowman & Sachs, 2008). Many of these lakes are important for the conservation of migratory waterbirds, including two of our study sites: Chaplin Lake and Old Wives / Frederick Lakes (Important Bird Areas, https://www.ibacanada.ca/).

We would expect that *Artemia* populations colonized the natural saline lakes of Saskatchewan following glacial retreat 11,000-15,000 years ago (Christiansen, 1979), with migratory waterbirds facilitating long-distance dispersal and hence colonization of newly formed lakes (Muñoz et al., 2013). Contemporary gene flow via bird movements may be expected to increase the similarity between lakes, particularly those that are close together, which may be visited repeatedly by the same individual birds even on a daily basis (Demers, Colwell, Takekawa, & Ackerman, 2008; Green, Figuerola, & Sánchez, 2002) and thereby increase the effect of IBD. Larger lakes are also likely to hold greater abundance and diversity of waterbirds (Bidwell, Green, & Clark, 2014; Sebastián-González & Green, 2014), increasing the arrival and diversifying the origin of cysts brought from other lakes, and hence promoting a correlation between lake size and genetic diversity. On the other hand, differences in hydrochemistry and resulting local adaptation may mean that only a subset of lakes provide suitable habitat for a given genotype, promoting IBE and reducing the chances that *Artemia* brought by birds as cysts may survive.

For this study, we apply the mitochondrial CO1 gene and microsatellite markers to study historic and current drivers of *Artemia franciscana* biogeography in its native distribution. In particular, we test the hypotheses that (1) genetic diversity is positively

126 correlated with lake area, and that (2) spatial genetic structure of populations in the
127 Saskatchewan saline lakes follows a pattern of isolation by environment (IBE) rather
128 than a pattern of isolation by distance that would result from dispersal limitation (IBD).

129 **Methods**

130 ***Study area***

131 From June to August 2011, 50 saline lakes were visited within the Saskatchewan plains
132 and searched for *Artemia*, although many of these lakes were too low in salinity to
133 support *Artemia franciscana*, and some of them only held other anostracans associated
134 with lower salinities (*Branchinecta* spp.). These lakes were partly selected on the basis
135 of existing literature on saline lakes, as some of them were included in broader studies
136 of the chemistry and invertebrate communities of saline lakes in Saskatchewan (Bentley,
137 Hayashi, Zimmerman, Holmden, & Kelley, 2016; Bowman & Sachs, 2008; Hammer, 1993;
138 Last & Ginn, 2005; Wissel, Cooper, Leavitt, & Pham, 2011). A minority of our study lakes
139 are well known for their *Artemia* populations, which have been used in previous
140 laboratory studies (e.g. Chaplin and Little Manitou; Bowen et al., 1988). Additional lakes
141 were sampled because they were encountered during ground surveys of Piping Plover
142 (*Charadrius melodus*) and other waterbirds for a separate study. Finally, other lakes
143 were visited because their saline nature was evident from inspection on Google Earth
144 (https://earth.google.com), with distinctive white soils around the water edge
145 indicating crystallization of saturated salts. Many of our study lakes are previously
146 undocumented, despite previous inventories of the worldwide distribution of *Artemia*
147 (Muñoz & Pacios, 2010). Our sampling was not exhaustive and additional unknown
148 populations of *Artemia* are likely to occur in Saskatchewan. Saline lakes of the Canadian
149 prairies are unaffected by commercial exploitation and export of *A. franciscana* cysts,

150 and the harsh climate makes aquaculture unviable, although adult *Artemia* have been
151 harvested in Chaplin Lake and some other lakes (Hammer, 1986).
152
153 ***Sampling***
154 Shorelines were visited during daylight by public access roads and the water column
155 was inspected for brine shrimps, especially on the side where winds concentrated them
156 along the shoreline by wave action. Where present, shrimps were collected with a hand
157 net then placed in alcohol for later confirmation of species identification. At the same
158 time, the shoreline was inspected for concentrations of cysts that form masses at the
159 water's edge or at the high water mark. These cysts were collected in 200 ml plastic
160 bottles, then later washed in distilled water before desiccation for better storage. A
161 water sample was also collected close to the shoreline in 200 ml plastic bottles which
162 were stored in a refrigerator for 1-6 weeks before they were analyzed for major ion
163 chemistry. Lake surface area was estimated from Google Earth (2012). Waterbirds were
164 present in all lakes at the time of visit (details available on request). Lakes used in this
165 study were separated from each other by up to 413 km.
166
167 ***Water chemistry analysis***
168 Alkalinity (Alk) was determined by sulfuric-acid titration, and concentrations of other
169 major ions were measured by ion-exchange chromatography after diluting brine
170 samples to a suitable concentration range.
171 Based on these measured parameters, we selected key descriptors of the variation
172 between lakes based on process-based understanding of hydrology and geochemistry of
173 saline lakes in the region (e.g., Bentley et al.(2016); Last & Ginn (2005)). These
174 descriptors were SO_4 concentration, and the ratios Cl/SO_4, Alk/SO_4, and Alk/(Ca +Mg).

Lake salinity is primarily controlled by the degree of evaporative enrichment of sulfate, which is sourced by the weathering of sulfide minerals (e.g. pyrite) contained in glacial till (Nachshon, Ireson, van der Kamp, & Wheater, 2013; Van Stempvoort, Hendry, Schoenau, & Krouse, 1994). Evaporative enrichment is greater in closed-basin lakes with higher water inputs by surface runoff and groundwater discharge (Wood & Sanford, 1990). Sulfate concentration is a useful indicator of enrichment to a point where the saturation of mirabilite ($Na_2SO_4 \cdot 10H_2O$) is reached. Further enrichment increases relative concentration of Mg compared to other cations. If the original composition of water has relatively high alkalinity, the geochemical evolution of water takes a path that is on the alkalinity side of the chemical divide (Hardie & Eugster, 1970). In contrast, originally alkalinity-poor water will take a path on the other side of the chemical divide characterized by high alkaline earth concentrations. Therefore, alkalinity/(Ca + Mg) ratio and alkalinity/SO_4 ratio may be useful for describing the geochemical evolution of saline lakes. Some saline lakes in Saskatchewan (including one in our study, PAT) receive brines from potash mines that contain high concentrations of chloride. Therefore Cl/SO_4 ratio is useful for distinguishing the effects of artificial inputs of chloride.

mtDNA genotyping

DNA was extracted from cysts sampled from 12 lakes (Table S1) in the study area using the Hotshot protocol (Montero-Pau, Gomez, & Muñoz, 2008). Two additional populations from two British Columbia lakes sampled in 2011 that belong to a lineage from the other side of the Continental Divide that includes populations of the Great Salt Lake, Utah, and San Francisco Bay (BCO and PCB, Table S1, Fig. S1) were included as outgroup in the haplotype network construction. We used *Artemia*-specific primers 1/2CO1_Fol-F and 1/2CO1_Fol-R

(Muñoz et al., 2008) to amplify a 709 bp fragment of the mitochondrial cytochrome *c* oxidase subunit 1 gene (CO1). Amplifications were performed in 20 µL total volume containing 1X reaction buffer, 2.0mM MgCl2, 0.2mM dNTPs, 0.6 units Taq DNA polymerase (Bioline Corp.) and 0.5µM of each primer. PCR conditions were as follows: 94ºC for 3 min, followed by 35 cycles of 45 s at 94ºC, 60 s at 45ºC (60-64ºC for 16S locus), and 60 s at 72ºC, followed by 5 min at 72 ºC. Quality of sequences was checked manually and consensus sequences were obtained and aligned in BioEdit v. 7.1.9 (Hall, 1999). The total number of newly obtained sequences varied between 16 and 24 per lake for 12 lakes. All corresponding haplotypes were deposited in NCBI GenBank (Accession numbers: MW799827-MW799878).

Nuclear microsatellite (nSSR) genotyping

DNA extractions were carried out on individual cysts (previously rinsed in distilled water) from each population making use of the HotSHOT protocol (Montero-Pau et al., 2008). The cysts were collected from *Artemia franciscana* populations of 14 lakes in Saskatchewan, Canada (Table 1, Table S1). We amplified eight nuclear microsatellite loci (in the following nSSR or simple sequence repeats) using the primer pairs described in Muñoz, Green, Figuerola, Amat, & Rico (2009) (Af_B109, Af_A108, Af_B9, Af_B139, Af_B117, Af_A136, Af_A104). The locus Af_B11 (Muñoz et al., 2009) was initially included but was dropped because its amplification failed in > 200 individuals. The results from the genotype accumulation curve computed with the R package *poppr* 2.8.3 (Kamvar, Tabima, & Grünwald, 2014) suggest that a full resolution of the multi-locus genotypes (MGLs) was already achieved by including only five loci (Fig. S2). We thus decided to use a higher number of genotyped individuals with the trade-off of removing locus Af_B11, yielding data from 592 individuals in total. Following tests for Hardy-

225 Weinberg equilibrium (HWE) (see below) we excluded two additional loci (Af_B117,
226 Af_A136) which deviated from HWE in all or almost all studied populations (Table S3) .
227 Therefore, throughout the paper, we report results for the five nSSR loci Af_B109,
228 Af_A108, Af_B9, Af_B139, Af_A104). All included loci were polymorphic and their allelic
229 richness was between 21 and 48 alleles (Table S4).
230
231 ***Statistical analysis***
232 All R packages used for data analyses described in this section were run in R version
233 3.6.2 (R Core Team, 2019).
234
235 *Environmental variables.* In order to quantify and visualize how the environment
236 available to *Artemia franciscana* differed between individual lakes, Principle Component
237 Analysis (PCA) was performed with R package *vegan* 2.5.6 (Oksanen et al., 2019) and
238 plotted with the autoplot() function of *ggplot2* (Wickham, 2016). The PCA included the
239 following variables: lake surface area, SO_4, and the ratios Cl/SO_4, Alk/SO_4, and Alk/(Ca
240 +Mg).

241 *mtDNA.* To analyze mtDNA, we supplemented the haplotype dataset newly sequenced in
242 this study, with haplotypes that were previously reported from four of our study lakes
243 (MAN, CHA, MEA and MUS), based on earlier sampling in 2009 (Muñoz et al., 2013). The
244 entire dataset used for the current study therefore included 65 haplotypes in total: 52
245 haplotypes newly detected in the present study, three haplotypes recorded both in the
246 present study and in Muñoz et al. (2013) and 10 additional haplotypes only reported in
247 Muñoz et al. (2013). Details on sample size and haplotypes can be found in Table S1 in
248 supplementary material.

A haplotype network was generated using TCS 1.21 (Clement, Posada, & Crandall, 2000) at the 95% connection limit. This network includes 12 lake populations from Saskatchewan and two lake populations from British Columbia. Genetic diversity within lakes was characterized by the number of private haplotypes (haplotypes found only in one location), standard diversity indices of haplotype diversity *h* and nucleotide diversity π (Nei 1987) calculated in Arlequin v. 3.5 (Excoffier & Lischer, 2010) and DnaSP v.5 (Librado & Rozas, 2009). Pairwise Φ_{ST} with Tamura & Nei (TrN) distances and their statistical significance (10,000 permutations) were performed in Arlequin v. 3.5.
An analysis of molecular variance (AMOVA) was conducted to determine the hierarchical structure of CO1 variation (Excoffier, Smouse, & Quattro, 1992). This analysis was performed with the aforementioned substitution model with 10,000 random permutations in Arlequin and based on the groups of populations that were not significantly genetically differentiated according to pairwise Φ_{ST} (Group 1: BEN, BES, WHI, FRE, CHA; Group 2: AKE; Group 3: EIN, LYD, MAN, MUS, PAT, MEA). We tested three hierarchical levels: between groups, between populations within groups, within populations.

Microsatellite (nSSR) markers. To estimate null allele frequencies, we used the software FreeNA (Chapuis & Estoup 2007) that follows the Expectation Maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) (detailed in Chapuis & Estoup 2007) . Hardy-Weinberg equilibrium was tested with the function hw.test(), computing an exact test based on 5000 Monte Carlo permutations of alleles. This function is implemented in the R package *pegas* 0.14 (Paradis 2010). Results were corrected for multiple comparisons using the method of Holm (1979), a less conservative modification of the Bonferroni test. Total allelic richness per populations was computed by rarefaction implemented in the R package *PopGenReport* 3.0.4 (Adamack & Gruber 2014). The population with the smallest number of individuals genotyped for nSSRs (SHO, 8

274 individuals) was excluded from the rarefaction procedures and the second smallest
275 population size (GUL, n = 21 individuals) applied as the smallest size for rarefaction.
276 Multilocus genotype diversity H (Shannon-Wiener diversity (Shannon, 1948)) was
277 computed with the R package *poppr* 2.8.3 (Kamvar et al., 2014). Locus- and population-
278 specific allele number and number of private alleles per population was calculated with
279 *poppr* 2.8.3 (Kamvar et al., 2014).
280 For pairwise population genetic distances, we computed Weir's (1996) F_{ST} using FreeNA
281 (Chapuis & Estoup 2007). This software provides a correction of the bias induced by null
282 alleles following the method described in Chapuis & Estoup (2007). We calculated
283 pairwise population D_{EST} with GenAlEx 6.502 (Peakall & Smouse, 2012) with associated
284 p-values based on 999 permutations.
285 To further explore the population structure suggested by pairwise population distance,
286 we applied discriminant analysis of principal components (DAPC) implemented in the
287 package adegenet 2.1.3 (Jombart 2008), a model-free analysis that makes no prior
288 assumptions about population genetic processes (such as HWE). We used the function
289 xvalDAPC() to find the optimum number of PCs to be retained prior to their use in the
290 final discriminant analysis, which we performed with the first 40 PCs and the first five
291 discriminant functions. We specified the 14 study populations as *a priori* groups to
292 analyse population membership probabilities for each individual.
293 *Distance-based redundancy analysis* (*dbRDA):* To test for the importance of isolation by
294 environment (IBE) vs. isolation by distance (IBD), we performed dbRDA with functions
295 implemented in the R package *vegan* 2.5.6 (Oksanen et al., 2019).
296 We used genetic distance estimators (pairwise D_{EST} for nSSR markers and pairwise Φ_{ST}
297 for mtDNA) as dependent variables. To test the effect of individual environmental
298 variables (the same ones as used in the above PCA) on population differentiation, we

299 built the null model (genetic differentiation ~ 1) and the full model (genetic
300 differentiation ~ Area + SO_4 + Cl/SO_4 + Alk/SO_4 + Alk/Ca+Mg) using the function
301 capscale(). Both the null model and the full model were used as input to the forward
302 selection model choice (on adjusted R^2 and p-values) with the function ordiR2step()
303 with 999 permutations. Model choice was repeated 100 times to obtain robust results.
304 To test the effect of spatial configuration, we computed distance-based Moran's
305 Eigenvector Maps (dbMEMs, previously known as PCNMs) using the R package
306 *adespatial* 0.3.7 (Dray et al., 2019) and chose the dbMEMs with associated Moran's I
307 values larger than expected under the null hypothesis of no spatial autocorrelation
308 (here: the first three dbMEMs: MEM1, MEM2, MEM3). For dbRDA, the null model was
309 tested against the full spatial model (genetic differentiation ~ MEM1 + MEM2 + MEM3)
310 using the forward selection model choice as described above.

311 *Network analysis.* To visualise the connectivity (as a proxy for gene flow) between
312 populations based on nSSR loci, we constructed a network graph using EDENnetworks
313 version 2.18 (Kivelä, Arnaud-Haond, & Saramäki, 2015) with manual thresholding,
314 based on F_{ST} corrected for null alleles obtained by FreeNA as decribed above. Manual
315 thresholding was performed at different values (0.02, 0.03 and 0.05) for visualisation of
316 possible population clusters under different threshold scenarios. Analysis of molecular
317 variance (AMOVA) for nSSR markers was used to test the significance of the population
318 structure suggested by the lowest threshold (0.02). Three hierarchical levels were
319 tested: (1) between 6 individual populations and two multi-population groups: AKE
320 (group1), WHI (group2), MUS (group3), MAN (group4), PAT (group5), FRE (group6),
321 BEN + BES (group7) and EIN + LYD + CHA + GUL + FRS + SHO (group8), (2) between
322 populations within groups, and (3) within populations. AMOVA was calculated using the

R package *ade4* 1.7-13 implemented in *poppr* 2.8.3 (Kamvar et al., 2014). Statistical significance of variance components was tested by permutation tests (999 permutations).

Results

Study sites – spatial distribution, lake size and hydrochemical properties

The 15 Saskatchewan saline lakes included in the genetic analysis spanned over a range of 2.63° in latitude, 3.93° in longitude, and were separated by distances of up to 410 km. Their spatial distribution was represented as Moran Eigenvectors for further analysis (Fig. S3). Their surface area ranged from 0.14 to 29.6 km^2 (Table 1). We obtained hydrochemical data for 13 of the lakes, including those for which nSSR data are available (with the exception of FRE and SHO, Table S2). The total dissolved solids of the 13 lake water samples ranged from 13 to 207 g L^{-1}, which were roughly proportional to SO_4 concentrations, except for PAT (see below).

A Principal Component Analysis (PCA) including SO_4 meq, the ratios Cl/SO_4, Alk/SO_4, Alk/(Ca+Mg) and lake area explained 68.9% of the environmental variation between lakes between the first two components (Fig. 1). Two lakes were clearly separated from the other sites: PAT was the site of a Potash mine and was associated with high values of Cl/SO_4, whereas AKE had particularly high Alk/(Ca+Mg). GUL was also hydrochemically separated from other lakes by a positive association with Alk/SO_4. There was a tendency for greater lake area to be associated with higher SO_4 (Fig. 1): MUS and WHI were among the largest lakes, and also had high SO_4.

347 *Haplotype and allelic diversity*

348 We identified a total of 52 new haplotypes in the sampled lakes. Of these, seven were
349 unique to the British Columbia locations. Three additional haplotypes were also
350 reported in a previous study, that also recorded 10 additional haplotypes not found in
351 our study (Muñoz et al., 2013). To construct the haplotype network and for all other
352 mtDNA analyses below, we included haplotype data from these authors (details in Table
353 S6).

354 Three star-like CO1 haplogroups formed by a total of 65 haplotypes exist in the
355 Saskatchewan area, each with several additional haplotypes of low frequency (Fig. 2).
356 These three main groups were centered on two haplotypes also recorded by Muñoz et
357 al. (2013) (Af_35, AF_41, Group A and C in Fig. 2) and a newly detected haplotype
358 (Hap_05, Group B). Overall, separation from the British Columbia haplotypes involved
359 about 4-8 mutations, while haplotypes within Saskatchewan were not strongly
360 differentiated (the two main haplotypes differed only by one mutation).

361 Diversity patterns differed between CO1 and nSSR markers (Table 1): CO1 haplotype
362 diversity ranged from 0.000 (PAT) to 0.779 (AKE). For nSSR, rarefied allelic richness
363 was between 26.3 (PAT) and 54.9 (FRS). Diversity of multilocus genotypes H ranged
364 from 2.493 (GUL) to 3.493 (EIN). No linear relationships were detected between allelic
365 or haplotype diversity and lake area (Table S7).

366

367 *Population genetic structure in geographic and environmental space*

368 There was significant structure between groups of populations according to the
369 mitochondrial CO1 gene (Table 2) with two main groups (BEN, BES, FRE, WHI, CHA, and
370 LYD, EIN, MAN, MUS, MEA, PAT), each of which was dominated by one of the two main
371 haplotypes. Differentiation between members of these two groups was strong, with

372 pairwise Φ_{ST} values between 0.548 and 0.891. AKE (with the greatest haplotype
373 diversity) was significantly differentiated from all other populations. An AMOVA of
374 mtDNA based on the above mentioned three groups confirmed that the variation was
375 highest between groups (68%), and much smaller between populations within groups
376 (1.37%). Within population variation was 31% (Table 3a).
377 While the pattern of differentiation was different from that of mitochondrial markers,
378 we also detected significant genetic differentiation between populations according to
379 pairwise D_{EST} based on nSSR loci (Table 2), ranged from -0.007 (between BEN and BES)
380 and 0.418 (between MUS and AKE).
381 Network visualisation based on F_{ST} corrected for null alleles (Table S5) revealed a
382 topology with two network components at a threshold of 0.02 (Fig. 3a). One included
383 EIN, LYD, GUL, CHA, FRS and SHO (some of which were well separated in space), while
384 the second included BES and BEN (which were closer together). The other five
385 populations remained disconnected at this threshold. A gradual increase of the
386 threshold led to higher complexity of the networks (Fig. 4b and c) with increasing
387 connectivity between the two clusters identified in Fig. 4a. However, AKE, WHI, PAT,
388 MAN and FRE remained separated even at the highest threshold. Using the two network
389 components identified at a threshold of 0.05, we found that genetic diversity (rarefied
390 allelic richness of nSSR loci) differed between connected and disconnected nodes (1-way
391 ANOVA: $F_{(1,11)} = 17.3$, $p = 0.002$) (Fig. 4d). This analysis suggested that populations
392 within a network component were genetically more diverse (median = 50.44, SD = 4.41)
393 than separated populations (median = 40.54, SD = 6.46).
394 The population groups identified from the network analysis at a threshold of 0.02 (Fig.
395 3a) were tested by AMOVA (Table 3b). The results were significant for all strata, but by
396 far the highest amount of variation was found within individual populations (83.50%).

Variation between groups of populations was 14.09%, suggesting a different pattern than that seen in mtDNA, as also suggested by the differences in group composition (e.g. the nearby BEN and BES group together with the more distant CHA and FRE by CO1, yet they form a separate group according to nSSR). Genetic variation of nSSR was lowest between populations within groups (6.09%).

To analyse environmental conditions and spatial configuration of sample sites that might underlie the observed genetic structure, we performed dbRDA using pairwise genetic distance (Φ_{ST} for the mitochondrial CO1 data, D_{EST} for nSSR loci, Table 2) as independent variables. No significant relationship was detected for mtDNA for either spatial or environmental variables. In contrast, for nSSR markers the forward selection model choice found strong and significant relationships with the ratio Alkalinity/(Ca +Mg) but not with spatial configuration. Alkalinity/(Ca +Mg) was significant in 87 of 100 repeats of the model choice ($p \leq 0.05$), explaining 30% of the observed genetic variation with a mean p-value of 0.04 (Table 4).

A discriminant analysis of principal components (DAPC, Fig. 4) of nSSR markers revealed a population genetic structure with six comparatively well-defined populations in which the majority of individuals were assigned to their respective population of origin (AKE, WHI, PAT, MUS, FRE and MAN), echoing the unconnected network nodes in Fig. 3. A considerable amount of admixture is suggested for other populations where EIN, LYD, GUL, CHA and FRS had higher membership probabilities for one of the westernmost populations while MUS, BEN and BES showed a high degree of admixture among themselves.

Discussion

As an important species for aquaculture, *Artemia franciscana* occurs widely outside its native distribution in North America. It is highly invasive, rapidly outcompeting native *Artemia* species. Invasive populations generally originate from commercially harvested cysts from the Great Salt Lake and San Francisco Bay, or secondary introductions from China (Amat et al., 2005; Horváth et al., 2018). In contrast, this species has rarely been studied in other areas of its native distribution, where it inhabits natural saline lakes with a wide range of hydrochemistries in particular regarding salt composition (Bowman & Sachs, 2008). In order to identify the roles of IBE and IBD in driving spatial genetic structure of *A. franciscana* in the saline lakes of Central Canada, we applied mitochondrial and nuclear markers to 14 lake populations in the Saskatchewan prairies. We found significant genetic structure between the Saskatchewan populations in both marker systems, and a strong relationship between the divergence pattern observed by nSSR markers and the hydrochemical environment of the lakes.

The mtDNA patterns observed are likely to reflect historical patterns of lake colonization, especially when lakes were formed after deglaciation, and then became saline enough to provide suitable habitat for *A. franciscana*, a species associated with higher salinities than all other anostracans in North America (Rogers, 2014a, 2014b). In the unique case of Patience (PAT), a lake contaminated with chloride by potash mining, colonization may have occurred in modern times after mining began. The presence of only one haplotype in PAT, combined with its unique pattern of nuclear markers, suggests strong and rapid local adaptation to these unusual chemical conditions absent from our other study sites, which were lakes in a more natural condition.

What mitochondrial markers reveal

For lakes other than PAT, our results suggest a close relationship of colonization by *Artemia* and deglaciation after the last glacial period, which first opened up habitat alongside the southwestern border of the ice sheet (14,000y BP). Under such a scenario, the area may first have been colonized in the western area by individuals originating from ice-free saline lake populations further to the west or south, such as Alberta or the USA. From the four western-most Saskatchewan populations (AKE, EIN, WHI, LYD), which hold all three haplotype groups identified in this study, dispersal may have proceeded towards the newly opened south-eastern habitats, subsequently colonized by one of the main haplotype groups (group C). Later, as the glacier retreated northeastwards and aquatic habitat opened towards the east in Saskatchewan, these northern lakes were largely colonized by the second main haplotype group (group A). Initial colonization by the two main haplotypes was followed by haplotype diversification, consistent with the observed star-like network pattern. After the area was ice-free in the next 1000 years (Christiansen, 1979), additional dispersal from the southern and western areas may have allowed colonization of these postglacial lakes. While this proposed sequence for colonization is consistent with the haplotype pattern observed in this study, it would require further study by including more *A. franciscana* populations from additional lakes in Saskatchewan and to the west in Alberta. Future sampling should also include populations further south in the Great Plains bioregion of the USA (Rogers, 2014b) to test the possibility that dispersal towards the northern areas may have followed the retreating ice sheet, and originated from regions south of Saskatchewan.

What nuclear markers reveal

469 In contrast, nSSR markers showed significant patterns of Isolation by Environment
470 (IBE), explaining ~30% of the variation in the genetic structure, but unrelated to
471 geographic configuration of the sampled lakes. The observed patterns are indicative of
472 local adaptation to the environment according to the criteria detailed in Orsini et al.
473 (2013). Since *Artemia* produces vast egg banks in established populations, ecological
474 and genetic priority effects occurring soon after initial colonisation and subsequent
475 monopolisation effects (De Meester, Gómez, Okamura, & Schwenk, 2002, De Meester,
476 Vanoverbeke, Kilsdonk, & Urban, 2016) may have reinforced such a pattern, including
477 the lack of a spatial signal that could be related to such priority effects. Our evidence for
478 IBE suggests that bird-mediated dispersal of cysts between lakes differing markedly in
479 hydrochemistry may not be effective (i.e. will not lead to colonization and establishment
480 in the new location). Laboratory tests show that nauplii larvae fail to reach adulthood
481 when placed in brine of unsuitable ionic composition (Bowen et al., 1988), and that cyst
482 hatching rates are sensitive to metal concentrations (Brix et al., 2006).
483 In several cases, strong differentiation is highly consistent with local adaptation to
484 extreme hydrochemical conditions, notably in PAT and AKE, but also in WHI, which had
485 the highest SO_4 concentrations in our study. In contrast to the mitochondrial data, nSSR
486 markers revealed that PAT was a highly divergent population, strongly related to an
487 elevated Cl/SO_4 ratio resulting from potash mining. Similarly, in AKE, this was related to
488 a particularly high Alkalinity/(Ca+Mg) ratio. Adaptation of different *A. franciscana*
489 populations to specific ion concentrations was demonstrated in the laboratory by
490 Bowens et al. (1988), leading to ecological isolation of individual populations and an
491 inability to colonize other lakes with divergent chemistry. This is also consistent with
492 the apparent influence of geochemistry on the evolution and distribution of different
493 anostracan taxa in North America (Rogers, 2014a). Since the nSSR markers are

494 essentially neutral, to the extent that lakes have similar genetic diversity due to similar
495 hydrochemistry, this cannot simply reflect convergent selection on genes for
496 osmoregulation or other key physiological processes. Instead, the genetic structure is
497 likely to reflect contemporary gene flow, which is inhibited by divergent chemistry but
498 promoted by a close spatial relationship between lakes.
499 Network analysis revealed a high connectivity between many populations, reflecting
500 their genetic similarity for nuclear markers. Together with PAT and AKE, a strongly
501 isolated population was FRE, which is surprising since this population is adjacent to FRS
502 and only isolated from it by a dam which also serves as a road. This creates two of
503 several separate sub-basins in Frederick Lake used for the extraction of sodium sulfate
504 via evaporation. The first basin (FRE) discharges water to FRS, which is a closed basin
505 with no surface water outlet. As a result, the water in FRE is expected to be less saline
506 compared to the concentrated brine in FRS, although we have no chemistry data for FRE.
507 Elsewhere, BES and BEN were also separated by a road, but were not differentiated
508 genetically, and had rather similar hydrochemistry, as they were parts of a larger lake
509 connected by culverts under the road.
510 Other lakes that were relatively isolated according to network analysis were WHI
511 (which was characterised by the highest SO_4 concentrations) and MAN. It is possible that
512 connectivity (i.e. genetic similarity) is influenced by the extent of interchange of
513 migratory birds between lakes. The most important site for waterbirds is CHA, which is
514 protected as a Ramsar site in recognition of the major concentrations of shorebirds it
515 holds. CHA was a central node in the network, connected with more populations in the
516 network than any other lake. On the other hand, adequate data on bird movements
517 between our study lakes do not exist, but our bird counts (AJ Green, unpublished data)
518 would suggest that WHI and MAN should be reasonably well-connected via waterbirds

519 with other lakes. WHI held thousands of Wilson's phalaropes *Phalaropus tricolor* which
520 were feeding on *Artemia* and are likely to be excellent vectors for their dispersal.
521 Hatching of nauplii from faeces collected at Frederick Lake from the American avocet
522 *Recurvirostra americana* during the course of this study confirmed the internal transport
523 of viable *A. franciscana* cysts in our study area (AJG, unpublished data). We did not find
524 evidence that lake area significantly affected genetic diversity. However, network
525 analysis showed that there is greater genetic diversity in highly connected populations
526 than in the less connected populations, and future research could test whether this
527 reflects movement patterns of birds between lakes. Outside of migratory periods,
528 waterbirds will generally provide higher connectivity for sites that are close together,
529 and hence coincide within the home range of more individual birds (Green et al., 2002).
530
531 *Conclusions*
532 Saline lakes of Saskatchewan provide a fascinating insight into the population genetics
533 of *Artemia* under natural conditions. The importance of IBE in this keystone crustacean
534 was only apparent as a consequence of our unusual application of hydrochemical
535 analysis. Studies of the population or invasion genetics of aquatic invertebrates in inland
536 waters typically limit their analyses of water quality to conductivity or salinity, and few
537 if any have gone into as much detail about ionic composition as in our study. More
538 frequent collaborations between hydrogeologists and molecular ecologists are
539 recommended. Although many of the sites occupied by other *Artemia* species in
540 different continents are coastal lagoons and salt ponds that are likely to have little
541 diversity in salt composition, many others are inland salt lakes (Muñoz & Pacios, 2010)
542 that may be hotspots of genetic diversity associated with local adaptation to unique
543 chemical conditions. On the other hand, the *Artemia* populations in such inland lakes

(e.g. *A. sinica* and *A. tibetiana* populations, Lin et al., 2017) may be resistant to invasion by *A. franciscana* populations originating from aquaculture, owing to contrasting hydrochemistry.

Acknowledgements

Robert G. Clark hosted AJG and provided vital assistance during sampling. Cysts from BC were kindly provided by Curtis A. Suttle, Amy M. Chan and Allyson L. Brady. Funding was provided by Spanish National Plan projects CGL2010-16028 and CGL2016-76067-P (AEI/FEDER, EU), and the PICS program funded by CSIC (PIC2015FR4) in Spain and CNRS (PICS07360) in France. DF received funding from the European Union's FP6-MOBILITY programme under the Marie Skłodowska Curie grant agreement No. 40285. AJG was also supported by mobility grants from the Spanish Ministry of Science and Innovation (PR2008-0293) and the Spanish Ministry of Education (PR2010-0602). Additional support was provided by Canadian Wildlife Service.

Data availability statement

Sequence data that support the findings of this study are available in NCBI GenBank, accession numbers MW799827-MW799878. Hydrochemical data is made available at digital.csic.es.

564 **References**

565 Adamack, A.T. & Gruber, B. (2014). PopGenReport: simplifying basic population genetic
566 analyses in R. *Methods in Ecology and Evolution*, 5(4), 384-387.

567 Amat, F., Hontoria, F., Ruiz, O., Green, A. J., Hortas, F., Figuerola, J., & Hortas, F. (2005).
568 The American brine shrimp as an exotic invasive species in the western
569 Mediterranean. *Biological Invasions*, 7(1), 37–47.

570 Bentley, L. R., Hayashi, M., Zimmerman, E. P., Holmden, C., & Kelley, L. I. (2016).
571 Geologically controlled bi-directional exchange of groundwater with a hypersaline
572 lake in the Canadian prairies. *Hydrogeology Journal*, 24(4), 877-892. doi:
573 10.1007/s10040-016-1368-0

574 Bidwell, M. T., Green, A. J., & Clark, R. G. (2014). Random placement models predict
575 species-area relationships in duck communities despite species aggregation. *Oikos,*
576 *123*(12), 1499-1508. doi: 10.1111/oik.00821

577 Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quarterly Review of*
578 *Biology*, *74*(1), 21–45.

579 Bowen, S. T., Buoncristiani, M. R., & Carl, J. R. (1988). *Artemia* habitats: Ion
580 concentrations tolerated by one superspecies. *Hydrobiologia, 158*(1), 201-214. doi:
581 10.1007/BF00026278

582 Bowman, J. S., & Sachs, J. P. (2008). Chemical and physical properties of some saline
583 lakes in Alberta and Saskatchewan. *Saline Systems,*4(1), 3. doi: 10.1186/1746-1448-
584 4-3

585 Brix, K. V, Gerdes, R. M., Adams, W. J., & Grosell, M. (2006). Effects of Copper, Cadmium,
586 and Zinc on the Hatching Success of Brine Shrimp (*Artemia franciscana*). *Archives of*
587 *Environmental Contamination and Toxicology*, *51*(4), 580–583.
588 https://doi.org/10.1007/s00244-005-0244-z

589 Chapuis, M.P., & Estoup, A. (2007). Microsatellite null alleles and estimation of
590 population differentiation. *Mol. Biol. Evol.* 24(3): 621-631.

591 Christiansen, E. A. (1979). Wisconsinan deglaciation of southern Sakatchewan and
592 adjacent areas. *Can J Earth Sci, 16(4)*, 913-938. doi: 10.1139/e79-079

593 Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate
594 gene genealogies. *Molecular Ecology 9(10),* 1657-1659. doi: 10.1046/j.1365-
595 294X.2000.01020.x

596 D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2018). Disentangling biotic
597 interactions, environmental filters, and dispersal limitation as drivers of species co-
598 occurrence. *Ecography, 41(8), 1233-1244*. doi: 10.1111/ecog.03148

599 De Meester, L., Gómez, A., Okamura, B., & Schwenk, K. (2002). The Monopolization
600 Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta*
601 *Oecologica*, *23*, 121–135.

602 De Meester, L., Vanoverbeke, J., Kilsdonk, L. J., & Urban, M. C. (2016). Evolving
603 Perspectives on Monopolization and Priority Effects. *Trends in Ecology and*
604 *Evolution*, 31(2),136-146. doi: 10.1016/j.tree.2015.12.009

605 Demers, S. A., Colwell, M. A., Takekawa, J. Y., & Ackerman, J. T. (2008). Breeding stage

influences space use of female American avocets in San Francisco Bay, California. *Waterbirds, 31(3), 365-371*. doi: 10.1675/1524-4695-31.3.365

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:1-38.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. H. (2019). adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-7.

Eimanifar, A., Marden, B., Braun, M. S., & Wink, M. (2015). Analysis of the genetic variability of *Artemia franciscana* Kellogg, 1906 from the Great Salt Lake (USA) based on mtDNA sequences, ISSR genomic fingerprinting and biometry. *Marine Biodiversity*, 45(2), 311-319. doi: 10.1007/s12526-014-0256-x

Eimanifar, A., Van Stappen, G., Marden, B., & Wink, M. (2014). *Artemia* biodiversity in Asia with the focus on the phylogeography of the introduced American species *Artemia franciscana* Kellogg, 1906. *Molecular Phylogenetics and Evolution*, 79, 392-403. doi: 10.1016/j.ympev.2014.06.027

Eimanifar, A., & Wink, M. (2013). Fine-scale population genetic structure in *Artemia urmiana* (Günther, 1890) based on mtDNA sequences and ISSR genomic fingerprinting. *Organisms Diversity and Evolution, 13(4), 531-543*. doi: 10.1007/s13127-013-0135-5

Excoffier, L, Smouse, P. E., & Quattro, J. M. (1992). Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes - Application to Human Mitochondrial-DNA Restriction Data. *Genetics*, *131*(2), 479–491.

Excoffier, Laurent, & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10(3), 564-567*. doi: 10.1111/j.1755-0998.2010.02847.x

Green, A. J., Figuerola, J., & Sánchez, M. I. (2002). Implications of waterbird ecology for the dispersal of aquatic organisms. *Acta Oecologica*, *23*, 177–189.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor. *Nucleic Acids Symposium Series*.

Hammer, U. Theodore. (1986). Saline Lake Resources of the Canadian Prairies. *Canadian Water Resources Journal*, *11*(1), 43–57. doi: 10.4296/cwrj1101043

Hammer, U.T. (1993). Zooplankton distribution and abundance in saline lakes of Alberta and Saskatchewan, Canada. *International Journal for Salt Lake Research*, *2*(111–132).

Hardie, L., & Eugster, H. (1970). Evolution of closed-basin brines. *Mineral. Soc. Amer. Spec. Pap. 3, 273-290.*

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Horváth, Z., Lejeusne, C., Amat, F., Sánchez-Fontenla, J., Vad, C. F., & Green, A. J. (2018). Eastern spread of the invasive *Artemia franciscana* in the Mediterranean Basin, with the first record from the Balkan Peninsula. *Hydrobiologia*. doi: 10.1007/s10750-

018-3683-z

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

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. doi: 10.7717/peerj.281

Kivelä, M., Arnaud-Haond, S., & Saramäki, J. (2015). EDENetworks: A user-friendly software to build and analyse networks in biogeography, ecology and population genetics. *Molecular Ecology Resources*, 15(1), 117-122. doi: 10.1111/1755-0998.12290

Last, W. M., & Ginn, F. M. (2005). Saline systems of the Great Plains of western Canada: an overview of the limnogeology and paleolimnology. *Saline Systems*,1,10. doi: 10.1186/1746-1448-1-10

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,25(11), 1451-1452*. doi: 10.1093/bioinformatics/btp187

Libralato, G., E. Prato, L. Migliore, A. M. Cicero, and L. Manfra. 2016. A review of toxicity testing protocols and endpoints with Artemia spp. Ecological Indicators **69**:35-49.

Lin, Q., Xu, L., Hou, J., Liu, Z., Jeppesen, E., & Han, B.-P. (2017). Responses of trophic structure and zooplankton community to salinity and temperature in Tibetan lakes: Implication for the effect of climate warming. *Water Research*, *124*, 618–629. https://doi.org/https://doi.org/10.1016/j.watres.2017.07.078

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

Muñoz, J., Amat, F., Green, A. J., Figuerola, J., & Gómez, A. (2013). Bird migratory flyways influence the phylogeography of the invasive brine shrimp *Artemia franciscana* in its native American range. *PeerJ, 1, e200*. doi: 10.7717/peerj.200

Muñoz, J., Gómez, A., Figuerola, J., Amat, F., Rico, C., & Green, A. J. (2014). Colonization and dispersal patterns of the invasive American brine shrimp *Artemia franciscana* (Branchiopoda: Anostraca) in the Mediterranean region. *Hydrobiologia, 726(1), 25-41*. doi: 10.1007/s10750-013-1748-6

Muñoz, J., Gomez, A., Green, A. J., Figuerola, J., Amat, F., & Rico, C. (2008). Phylogeography and local endemism of the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda : Anostraca). *Molecular Ecology*, *17*(13), 3160–3177.

Muñoz, J., Green, A. J., Figuerola, J., Amat, F., & Rico, C. (2009). Characterization of polymorphic microsatellite markers in the brine shrimp *Artemia* (Branchiopoda, Anostraca). *Molecular Ecology Resources*, *9*(2), 547–550.

Muñoz, J., & Pacios, F. (2010). Global biodiversity and geographical distribution of diapausing aquatic invertebrates: The case of the cosmopolitan brine shrimp, *Artemia* (Branchiopoda, Anostraca). *Crustaceana*, 83(4), 465-20.A. doi: 10.1163/001121610X489449

690 Nachshon, U., Ireson, A., van der Kamp, G., & Wheater, H. (2013). Sulfate salt dynamics in
691 the glaciated plains of North America. *Journal of Hydrology, 499, 188-199*. doi:
692 10.1016/j.jhydrol.2013.07.001

693 Naihong, X., Audenaert, E., Vanoverbeke, J., Brendonck, L., Sorgeloos, P., & De Meester, L.
694 (2000). Low among-population genetic differentiation in Chinese bisexual *Artemia*
695 populations. *Heredity, 84(2), 238-243*. doi: 10.1046/j.1365-2540.2000.00664.x

696 Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation
697 between walking-stick ecotypes: 'Isolation by adaptation' and multiple roles for
698 divergent selection. *Evolution, 62(2), 316-336*. doi: 10.1111/j.1558-
699 5646.2007.00299.x

700 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H.
701 (2019). vegan: Community Ecology Package. R package version 2.5-6.

702 Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of
703 population genetic differentiation in the wild: Isolation by dispersal limitation,
704 isolation by adaptation and isolation by colonization. *Molecular Ecology, 22(24),*
705 *5983-5999*. doi: 10.1111/mec.12561

706 Ovaskainen, O., Tikhonov, G., Dunson, D., Grotan, V., Engen, S., Saether, B. E., & Abrego, N.
707 (2017). How are species interactions structured in species-rich communities? A
708 new method for analysing time-series data. *Proc Biol Sci*, *284*(1855), 20170768. doi:
709 10.1098/rspb.2017.0768

710 Paradis, E. (2010). pegas: an R package for population genetics with an integrated–
711 modular approach. *Bioinformatics*, 26, 419-420.

712 Peakall, R., & Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population
713 genetic software for teaching and research-an update. *Bioinformatics, 28(19), 2537-*
714 *2539*. doi: 10.1093/bioinformatics/bts460

715 R Core Team (2019). (2019). R: A language and environment for statistical computing.
716 *Accessed 1st April 2019*.

717 Reynolds, C., Miranda, N. A. F., & Cumming, G. S. (2015). The role of waterbirds in the
718 dispersal of aquatic alien and invasive species. *Diversity and Distributions, 21(7),*
719 *744-754*. doi: 10.1111/ddi.12334

720 Rogers, D. C. (2014a). Anostracan (Crustacea: Branchiopoda) zoogeography II. Relating
721 distribution to geochemical substrate properties in the USA. *Zootaxa*, 3838(3), 1.
722 doi: 10.11646/zootaxa.3856.1.1

723 Rogers, D. C. (2014b). Anostracan (Crustacea: Branchiopoda) zoogeography I. North
724 American bioregions. *Zootaxa*, 3838(3), 251. doi: 10.11646/zootaxa.

725 Ruebhart, D. R., I. E. Cock, and G. R. Shaw. 2008. Brine shrimp bioassay: Importance of
726 correct taxonomic identification of Artemia (Anostraca) species. Environmental
727 Toxicology **23**:555-560.

728 Sánchez, M. I., Hortas, F., Figuerola, J., & Green, A. J. (2012). Comparing the potential for
729 dispersal via waterbirds of a native and an invasive brine shrimp. *Freshwater*
730 *Biology*, 57(9), 1896-1903. doi: 10.1111/j.1365-2427.2012.02852.x

731 Sebastián-González, E., & Green, A. J. (2014). Habitat use by waterbirds in relation to

pond size, water depth, and isolation: Lessons from a restoration in southern Spain. *Restoration Ecology, 22(3), 311-318*. doi: 10.1111/rec.12078

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal 27(3), 379-423*. doi: 10.1002/j.1538-7305.1948.tb01338.x

Van Stempvoort, D. R., Hendry, M. J., Schoenau, J. J., & Krouse, H. R. (1994). Sources and dynamics of sulfur in weathered till, Western Glaciated Plains of North America. *Chemical Geology, 111(1-4), 35-56*. doi: 10.1016/0009-2541(94)90081-7

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. doi: 10.1111/mec.12938

Weir, B. S. (1996). Genetic Data Analysis II. Sinauer Associates, Sunderland, Mass.

Wickham, H. (2016). ggplot2 Elegant Graphics for Data Analysis (Use R!). In *Springer*. doi: 10.1007/978-0-387-98141-3

Wissel, B., Cooper, R. N., Leavitt, P. R., & Pham, S. V. (2011). Hierarchical regulation of pelagic invertebrates in lakes of the northern Great Plains: A novel model for interdecadal effects of future climate change on lakes. *Global Change Biology, 17(1), 172-185*. doi: 10.1111/j.1365-2486.2010.02291.x

Wood, W. W., & Sanford, W. E. (1990). Ground-water control of evaporite deposition. *Economic Geology, 85(6), 1226-1235*. doi: 10.2113/gsecongeo.85.6.1226

Wright, S. (1943). Isolation by distance. Genetics. *Genetics*, *28*(28), 114–138.

Figure captions

Fig. 1 Principal Component Analysis of environmental variables in Saskatchewan saline lakes. The percentage of total variation explained by the first two components is given on the respective axes.

Fig. 2. a) Haplotype network (mtDNA) for Canadian *Artemia franciscana* (includes CO1 haplotypes from Saskatchewan and British Columbia). Grey squares highlight the four main haplotype groups present in Saskatchewan. The top nodes of the network (grey shades) represent members of two populations (BCO, PBC) in British Columbia ~1000 km west of the Saskatchewan study area. Haplotype groups A–C are based on genetic differentiation between the populations that hold these haplotypes as identified by pairwise ΦST and subsequent AMOVA (see Tables 2,3).
b) Geographic distribution of haplotype groups A–C. Numbered arrows indicate the possible sequence of colonization after deglaciation (details in the text). Lines indicate the glacial border at 14,000y BP (blue), 12,500y BP (green, dotted) and 11,500y BP (red, hatched) (drawn after Christiansen, 1979). Scale is approximate, for exact distances between lakes see Table S8. Note: to avoid confusion between patterns of mitochondrial and nuclear markers, we applied different colour schemes in this figure and Fig. 4.

Fig. 3 a -d. EDENetwork for *Artemia franciscana* in Saskatchewan constructed from nSSR loci. Each population is positioned according to its geographical coordinates. Edges (connections between populations) are shown for three different thresholds at (a) 0.02,

(b) 0.03 and (c) 0.05, thus successively allowing weaker connections to connect nodes, adding complexity to the network The thickness of edges is proportional to the genetic distance between populations, (thicker lines denoting less distance) while the size of nodes is relative to the number of connections between nodes. Green nodes have at least one connection, whereas red nodes are unconnected. (d) box plot (with median, comparing the genetic diversity (nSSR, AR = rarefied allelic richness) of connected (y) and isolated (n) populations (connections as in (a)). Dots represent measurements for each population, with outliers in red.

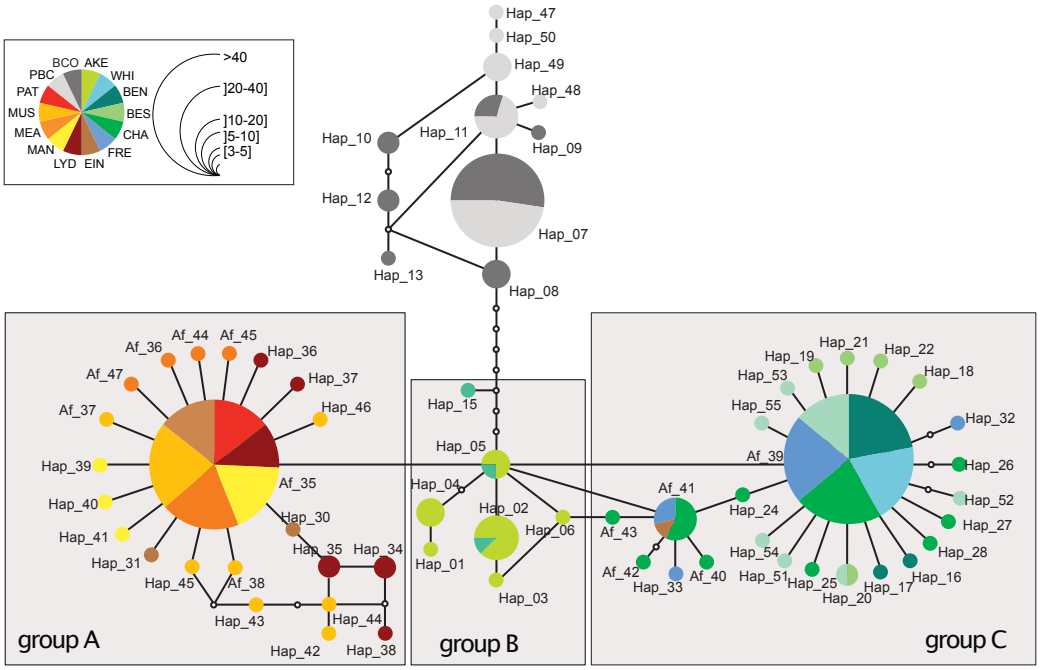
Fig. 4 a-b. a) Population genetic structure of *Artemia franciscana* according to DAPC based on nuclear markers (nSSR, see Methods for details). Each individual is represented by a vertical bar, coloured by estimated membership to one of the sampled populations. Posterior membership probability < 1.0 results in multiple assignments (and thus multi-coloured bars) per individual.. Bars below the graph represent the dominant CO1 haplotype groups (A, B, and C, from Fig. 2) detected in each population (not available for GUL, FRE, SHO, denoted by "?"). Note: to avoid confusion between patterns of mitochondrial and nuclear markers, we applied different colour schemes in this figure and Fig. 2.
b) Map with sample locations and representation of the percentage of individuals from each lake assigned to each of the 14 populations (by DAPC). Scale is approximate, for exact distances between lakes see Table S8. Colour legend for both figures in 4b)

Fig. S1. Phylogenetic relationships of native Artemia franciscana COI haplotypes after Muñoz et al. [1] with two populations added from British Columbia, Canada, that were sampled in the present study (BCO, PBC). The evolutionary history was inferred using the Neighbor-Joining method. Results of bootstrap test (1000 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [3] and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7 [4]. Each lineage label (redrawn from [1]) indicates which countries it is found in and its overlap with the Pacific, Atlantic or Central migratory flyways for shorebirds (P, A or C respectively)
Fig. S2 Genotype accumulation curve

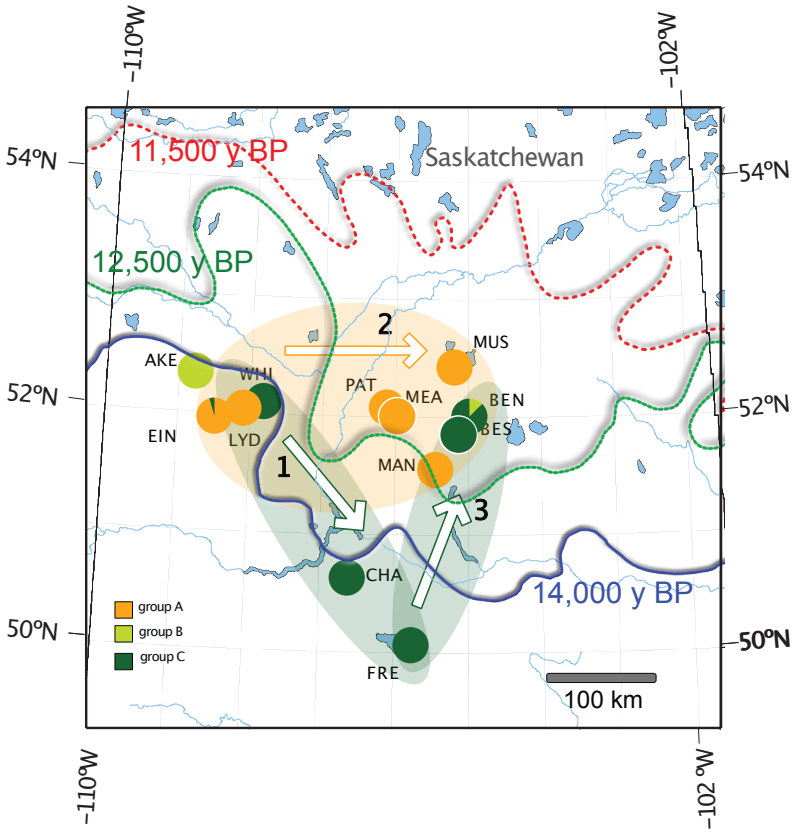
Fig. S3. Graphical representation of MEM1, MEM2 and MEM3, the distance based Moran Eigenvector Maps used as spatial predictors in the dbRDA. Size of circles represents absolute eigenvalues of each dbMEM (positive values: filled circles, negative values: open circles). Circles are positioned according to the latitude and longitude of individual saline lakes in Saskatchewan.

Fig. S4. Principal Coordinate analysis (PCoA), based on Jost's D_{est}. Populations are represented by STRUCTURE population clusters (K=3, based on nSSR markers See Fig. 3a for details).

a)

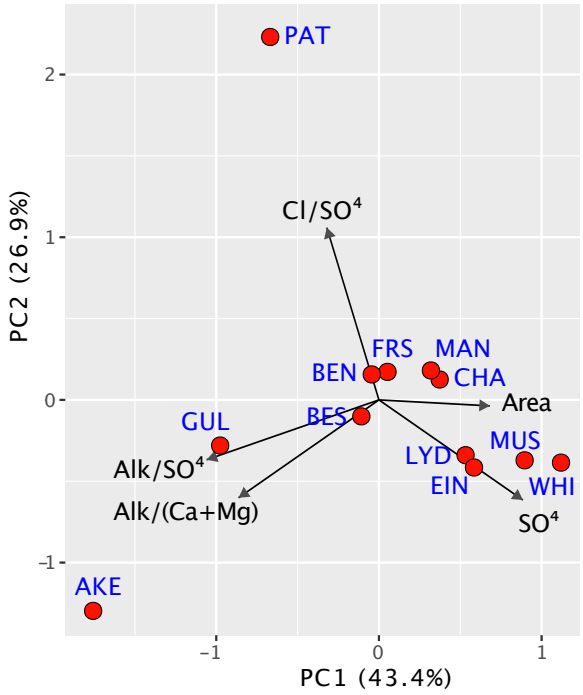

BCO AKE
PBC
WHI
PAT
BEN
MUS
BES
MEA
CHA
MAN
FRE
LYD EIN
>40
]20-40]
]10-20]
]5-10]
[3-5]
Hap_47
Hap_50
Hap_49
Hap_48
Hap_10
Hap_11
Hap_09
Hap_12
Hap_07
Hap_13
Hap_08
Hap_15
Hap_05
Hap_04
Hap_02
Hap_06
Hap_01
Hap_03
Af_41
Af_43
Af_42
Af_40
Hap_33
Hap_24
Af_39
Hap_21
Hap_22
Hap_19
Hap_18
Hap_53
Hap_55
Hap_32
Hap_26
Hap_52
Hap_27
Hap_28
Hap_54
Hap_51
Hap_25
Hap_17
Hap_16
Hap_20
Af_36
Af_44
Af_45
Hap_36
Af_47
Hap_37
Af_37
Hap_46
Hap_39
Af_35
Hap_40
Hap_30
Hap_41
Hap_35
Hap_34
Hap_31
Hap_45
Af_38
Hap_43
Hap_44
Hap_42
Hap_38
group A
group B
group C

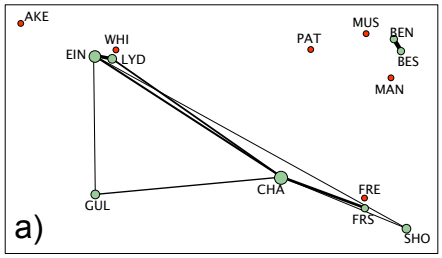
b)

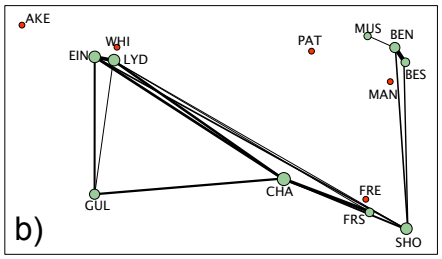

-110°W
-102°W
54°N
52°N
50°N
11,500 y BP
12,500 y BP
Saskatchewan
14,000 y BP
2
1
3
MUS
AKE
WHI
PAT
MEA
BEN
BES
EIN
LYD
MAN
CHA
FRE
group A
group B
group C
100 km

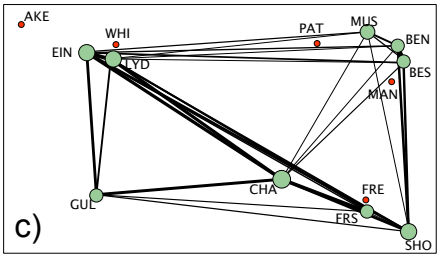
820

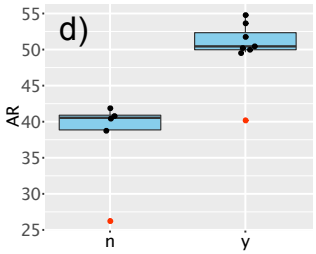
821


PAT
Cl/SO4
FRS
MAN
BEN
CHA
Area
BES
GUL
MUS
LYD
Alk/SO4
EIN
WHI
Alk/(Ca+Mg)
SO4
AKE
PC2 (26.9%)
PC1 (43.4%)
2
1
0
−1
−1
0
1

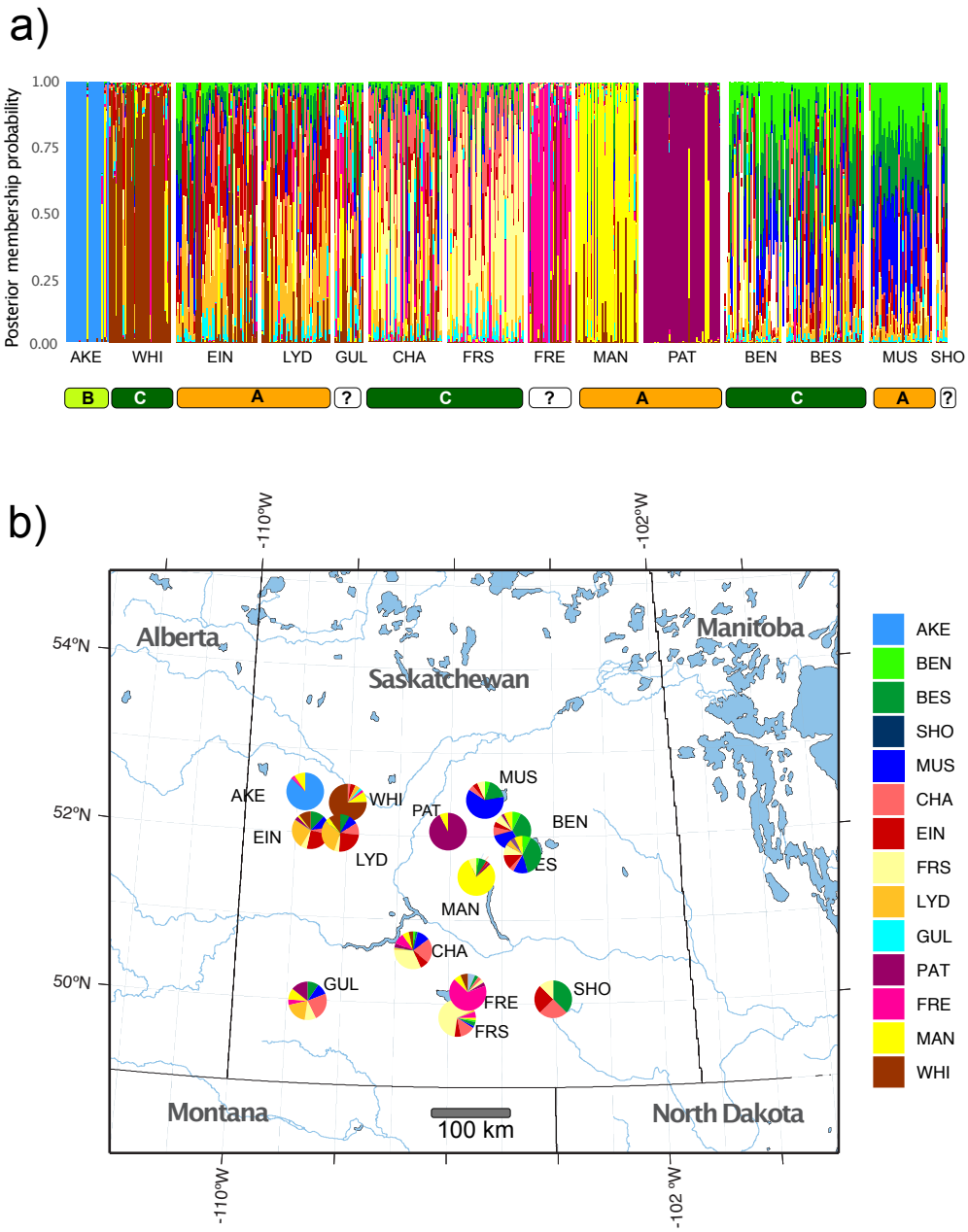

AKE
WHI
EIN
LYD
PAT
MUS
BEN
BES
MAN
CHA
FRE
GUL
FRS
SHO
a)


AKE
WHI
EIN
LYD
PAT
MUS
BEN
BES
MAN
CHA
FRE
GUL
FRS
SHO
b)


AKE
WHI
EIN
LYD
PAT
MUS
BEN
BES
MAN
CHA
FRE
GUL
FRS
SHO
c)


d)
AR
55
50
45
40
35
30
25
n
y

822



823
824
825 *Authors' contribution statement*
826
827 AJG conceived the study and collected samples. MTB collected samples. CL and JSF

generated the molecular data. MH generated and analysed the hydrochemical data. DF analysed the molecular data with CL and wrote the paper with AJG and input from all other authors.

Conflict of interest

The authors declare no conflict of interest.

834 TABLES

835 Table 1. Genetic diversity of 14 *Artemia franciscana* populations from Saskatchewan. For
836 nSSR markers (5 loci), rarefied total allelic richness (AR) is indicated as well as the
837 number of private alleles (A_P) and the Shannon-Wiener index for multilocus genotypes
838 (H). SHO is excluded here, as only 8 individuals were genotyped. For mtDNA, haplotype
839 diversity (*h)*, nucleotide diversity (π) and the number of private haplotypes (h_p) are
840 indicated. Lake size is given as surface area in km^2.
841

	nSSR			**mtDNA**			
Pop	**AR**	**A_P**	**H**	***h***	**π**	**h_p**	**Area (km^2)**
AKE	40.9	7	2.876	0.779	0.00290	4	3.282
MAN	40.5	2	3.242	0.270	0.00047	4	15.599
PAT	26.3	4	3.445	0.000	0.00000	0	5.725
WHI	42.0	5	3.240	0.569	0.00129	5	29.618
BEN	50.6	4	3.149	0.380	0.00134	3	0.140
BES	53.8	1	3.442	0.411	0.00075	4	0.707
CHA	49.6	8	3.405	0.641	0.00246	8	23.206
EIN	50.3	4	3.493	0.260	0.00060	2	1.964
FRE	38.9	3	2.873	0.320	0.00122	2	0.890
FRS	54.9	9	3.441	NA	NA	NA	0.590
LYD	51.9	5	3.324	0.537	0.00223	5	2.061
GUL	50.1	4	2.493	NA	NA	NA	0.576
MUS	40.3	0	3.243	0.356	0.00123	7	18.284
MEA	NA	NA	NA	0.253	NA	4	0.900

Table 2: Pairwise genetic distances between *Artemia franciscana* populations in Saskatchewan. The lower diagonal shows Pairwise Φ_{ST} (Tamura & Nei) based on the mtDNA marker (CO1). The upper diagonal shows D_{EST} computed for five nSSR markers. Numbers in light grey have a non-significant associated p-value (> 0.01).

	AKE	**BEN**	**BES**	**FRE**	**FRS**	**WHI**	**CHA**	**LYD**	**EIN**	**MEA**	**MAN**	**MUS**	**PAT**	**SHO**	**GUL**
AKE		0.341	0.321	0.326	0.374	0.316	0.316	0.332	0.296	-	0.323	0.418	0.331	0.322	0.313
BEN	0.476		-0.007	0.318	0.150	0.334	0.077	0.093	0.069	-	0.253	0.048	0.204	0.038	0.159
BES	0.574	0.016		0.294	0.155	0.296	0.075	0.086	0.058	-	0.235	0.070	0.181	0.044	0.139
FRE	0.493	0.000	0.038		0.188	0.140	0.159	0.233	0.213	-	0.108	0.391	0.225	0.260	0.166
FRS	-	-	-	-		0.305	0.015	0.072	0.086	-	0.255	0.139	0.326	0.082	0.105
WHI	0.521	0.013	-0.006	0.025	-		0.219	0.182	0.159	-	0.100	0.374	0.215	0.293	0.157
CHA	0.386	0.061	0.105	0.009	-	0.090		0.029	0.024	-	0.178	0.089	0.220	0.022	0.038
LYD	0.490	0.636	0.700	0.643	-	0.658	0.548		0.004	-	0.213	0.097	0.258	0.058	0.075
EIN	0.570	0.736	0.824	0.750	-	0.777	0.593	0.085		-	0.168	0.082	0.192	0.034	0.046
MEA	0.622	0.764	0.839	0.777	-	0.801	0.637	0.123	0.001		-	-	-	-	-
MAN	0.623	0.769	0.848	0.783	-	0.808	0.637	0.123	0.001	0.000		0.310	0.149	0.251	0.121
MUS	0.568	0.698	0.761	0.706	-	0.728	0.602	0.048	0.015	0.029	0.028		0.295	0.076	0.152
PAT	0.623	0.794	0.891	0.813	-	0.841	0.632	0.115	-0.007	-0.016	-0.015	0.016		0.229	0.218
SHO	-	-	-	-	-	-	-	-	-	-	-	-	-		0.081

Table 3 a, b. Population genetic structure of *Artemia franciscana* in Saskatchewan identified by AMOVA for (a) CO1 haplotypes (mtDNA) with p-values obtained from 10,100 randomisations as implemented in Arlequin v3.5.2.2, and (b) five nSSR loci, with associated p-values obtained by Monte Carlo permutation test with 999 repeats. For a), population genetic structure was tested for the following groups of populations with low genetic differentiation according to Table 2 (lower diagonal). Group1: AKE; group2: BEN, BES, FRE, WHI, CHA; group3: EIN, LYD, MUS, MAN, PAT, MEA. Groups of populations for b) were chosen according to connected and unconnected populations resulting from network construction (Fig. 3a): AKE (group1), WHI (group2), MUS (group3), MAN (group4), PAT (group5), FRE (group6), group7: BEN, BES, and group8: EIN, LYD, CHA, GUL, FRS, SHO.

,

Source of variation	**d.f.**	**Sum of squares**	**Variance components**	**% variation**	**p-value**
a) CO1 haplotypes					
Between groups	2	135.42	0.839	68.06	$<10^{-5}$
Between populations within groups	9	7.18	0.017	1.37	$<10^{-5}$
Within populations	281	105.96	0.377	30.57	$<10^{-5}$
Total	292	248.56	1.233		
b) nSSRs					
Between groups	7	176.90	0.321	14.09	0.001
Between populations within groups	6	24.70	0.055	2.41	0.001
Within populations	578	1100.03	1.903	83.50	0.001
Total	591	1301.62	2.279		

Table 4. Summary of results of dbRDA for nSSR markers obtained by forward model selection (100 repeats) to test the effect of individual environmental variables (lake area, SO_4, Cl/SO_4, $Alkalinity/SO_4$, Alkalinity/Ca+Mg) on genetic distance (D_{EST}, for details see Methods). Lake area, SO_4, Cl/SO_4, $Alkalinity/SO_4$ were not selected by forward model selection and therefore were not included in the table. "*signif* (n)" = the number of times that results were significant (of 100 repeats).

statistics	Alk/Ca+Mg
R2.adj	0.30
AIC	-16.3
F	4.49
p.mean	0.04
p.min	0.024
p.max	0.05
p.SD	0.007
signif (n)	87