

The invasion history of the exotic freshwater zooplankter *Daphnia lumholtzi* (Cladocera, Crustacea) in North America: a genetic analysis

Dagmar Frisch · John E. Havel ·
Lawrence J. Weider

Abstract *Daphnia lumholtzi* is a planktonic crustacean native to subtropical regions in Africa, Asia and Australia. Since its invasion to the southern USA in ~1990 it has spread across North America as far north as the Laurentian Great Lakes. We assessed invasion history using microsatellite markers and to explore the influence of mean annual temperature on the genetic structure along a latitudinal gradient in North America. Genotypic data were obtained from 9 microsatellite markers for 178 individuals from 13 populations (eight populations introduced to North America and

five populations in the native range). Pairwise F_{st} values as well as Bayesian clustering showed a strong subdivision between native and introduced populations. Bayesian clustering identified multiple genetic clusters in recently invaded locations, suggestive of multiple invasions from various sources, including Asia and Africa. Using variation partitioning, we determined the amount of variation for genetic clusters of populations in the invaded range due to mean annual air temperature and the year of first detection. The results point to a primary introduction into the southern range of North America, with a subsequent northward expansion, and multiple introductions possibly from both the native range and by secondary spread from previously-invaded locations. Separate analysis of genetic clusters within the invaded range suggests additional effects of temperature conditions on geographic genetic structure, possibly as a consequence of *D. lumholtzi*'s tropical origin.

Electronic supplementary material The online version of this article (doi:[10.1007/s10530-012-0329-3](https://doi.org/10.1007/s10530-012-0329-3)) contains supplementary material, which is available to authorized users.

D. Frisch (✉) · L. J. Weider
University of Oklahoma Biological Station, 15389 Station
Road, Kingston, OK 73439, USA
e-mail: dfrisch@sistern.net

D. Frisch
Doñana Biological Station EBD-CSIC, c/Americo
Vespucio, s/n, 41092 Sevilla, Spain

J. E. Havel
Department of Biology, Missouri State University,
901 South National, Springfield, MO 65897, USA

L. J. Weider
Department of Biology, Program in Ecology
and Evolutionary Biology, University of Oklahoma,
730 Van Vleet Oval, Norman, OK 73019, USA

Keywords Biological invasion · Macrogeographic genetic structure · Range expansion · Zooplankton · Population genetics · Microsatellites

Introduction

Global translocations of freshwater zooplankton offer an interesting model for understanding the importance of history and climate on macrogeographic population structure. In North America, a variety of zooplankton have invaded lakes, many carried long distances in the

ballast water of large ships (MacIsaac et al. 1999; Cristescu et al. 2001; Havel and Medley 2006). The best-studied example, the spiny water flea *Bythotrephes longimanus* (Leydig, 1860), has neither spread far beyond the primary invasion area nor colonized across various climate zones of North America (Yan et al. 2011; and <http://nas.er.usgs.gov>). In contrast, *Daphnia lumholtzi* Sars, 1885 a native to the subtropics of Africa, Asia, and Australia, was likely transported to North America with introduced fish (Sorensen and Sterner 1992) and quickly spread across the United States (Havel and Shurin 2004). Now common across a wide latitudinal (i.e. temperature) range from Florida to the Great Lakes, *D. lumholtzi* is capable of being dispersed via anthropogenic vectors (i.e. recreational boats—Havel and Stelzleni-Schwent 2000). Thus, the invasion of *D. lumholtzi* is an exceptional case of a zooplankton invasion with regard to its initial route of introduction to North America, its speed of secondary spread, and the geographical and climatic range invaded. Since the year 2000, the species has also been recorded in three sites in Brazil (Zanata et al. 2003; Simoes et al. 2009). *D. lumholtzi* has a conspicuous morphology that is clearly distinct from *Daphnia* species native to North America (Havel and Hebert 1993), and is thus readily identifiable in water samples. Havel and Hebert (1993) suggested African populations as a possible introductory source for North American populations, and considered the commercial stocking of Nile Perch from Lake Victoria in reservoirs of the southern US in the 1980s as the most likely introduction pathway. Results from allozymes and mitochondrial (mt)DNA suggested a close relationship of North American populations with African and Asian populations, but not with Australian populations (Havel et al. 2000).

While direct observations of species invasions provide important records on invasion extent and history (Havel et al. 2002), genetic data for invasive species have yielded key information on the geographical origin of source populations (Ascunce et al. 2011). The study of population genetic structure of invasives offers insights about origins and the number of introductions during the invasion process (Reusch et al. 2010). Such knowledge on geographic origin and number of introductions guides our understanding of invasion trajectories, providing insight into management strategies for reducing the chance of future introductions (Estoup and Guillemaud 2010; Fitzpatrick et al. 2012).

Genetic diversity has sometimes been linked to population age, but the relationship between these variables is not altogether clear. Loss of genetic diversity has been observed in recently established populations as a result of founder effects (Nei et al. 1975), strong selection, and genetic drift in the novel environment (reviewed in Dlugosch and Parker 2008). Repeated introductions can contribute to a positive relationship between genetic diversity and population age in invasive populations, such that older populations are more diverse than younger (Dlugosch and Parker 2008). The opposite pattern, i.e. higher genetic diversity in younger populations, can result from high propagule pressure and multiple invasions from different geographic sources (Roman and Darling 2007) and may lead to genetic diversity equal to that of native populations (Kelly et al. 2006). For example, younger populations of South African ragwort (*Senecio inaequidens*) were more diverse than older populations due to sufficient gene flow among populations along invasion routes (Lachmuth et al. 2010).

Here we present the first study to use highly variable molecular markers (i.e. microsatellites) for assessing the invasion history of *D. lumholtzi*. We specifically aim to explore the population genetic structure along latitudinal gradients with mean annual air temperature as a proxy for different climatic conditions across North America, and to discriminate between a single invasion with subsequent range expansion versus multiple introductions. We test the following hypotheses: (1) The *D. lumholtzi* invasion in North America constitutes a single source introduction in the southern USA with subsequent range expansion by secondary introductions to new sites. (2) There is a latitudinal population genetic structure of *D. lumholtzi* along a temperature gradient within the invaded range of North America.

Methods

Sample collection and preservation

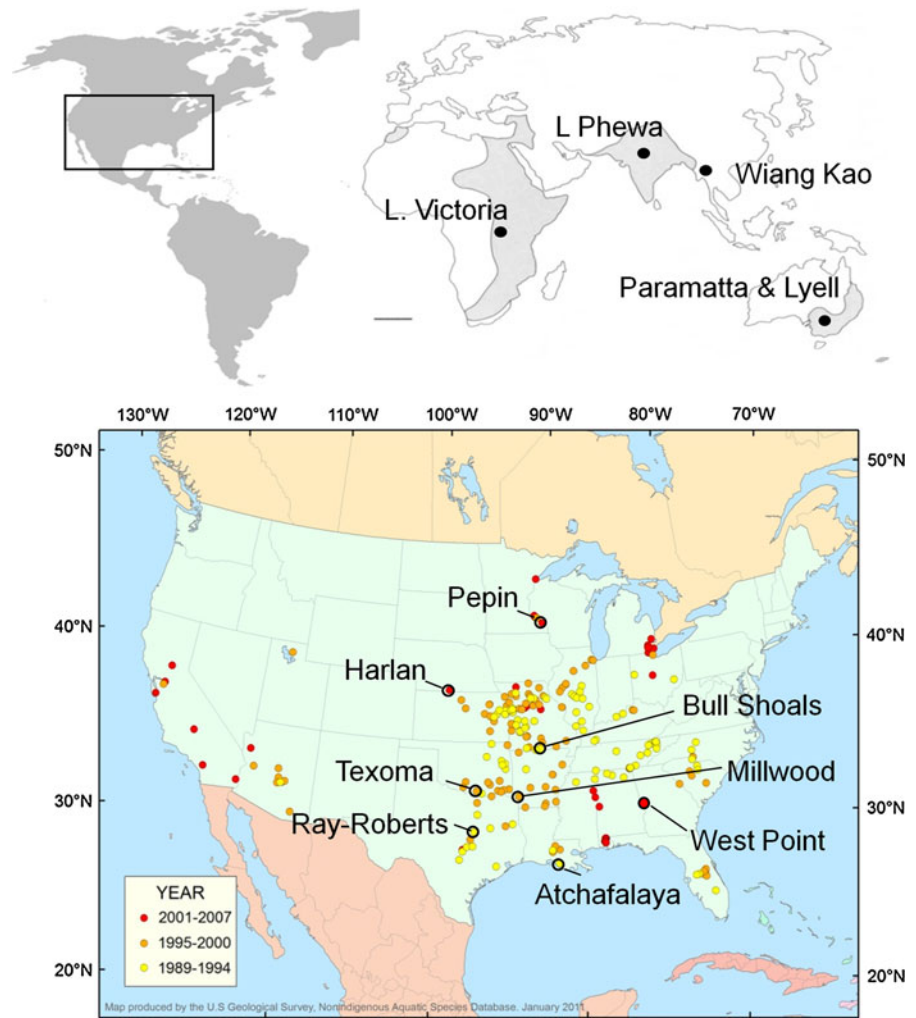
Daphnia lumholtzi was collected from eight North American populations and five populations throughout its native range (geographical details in Table 1; Fig. 1). Samples were preserved either in a DMSO (dimethyl-sulfoxide) solution containing 20 % DMSO, 0.25 M sodium-EDTA, and NaCl to

Table 1 Summary of population genetic diversity in microsatellite markers of native and non-native populations of *D. lumholtzi*

Site	first record (year)	<i>N</i>	Loci									
			Dma1	B008	S6-199	B064	Dm14	B096	Dma15	Dma3R	Dma11	across loci
∑ all populations		<i>N_a</i>	5	5	17	5	8	5	11	10	5	-
Pepin, MN 44.4102° -92.1003°	1999	<i>N</i>	14	13	5	8	3	4	10	3	9	7.7
		<i>N_a</i>	4	3	5	1	1	3	3	1	1	2.4
		<i>H_o</i>	0.429	0.462	0.200	0.000	0.000	0.250	0.100	0.000	0.000	0.160
		<i>H_e</i>	0.441	0.601	0.780	0.000	0.000	0.531	0.185	0.000	0.000	0.282
West Point, GA 32.9563° -85.1925°	2007	<i>N</i>	12	13	8	11	1	4	8	3	4	7.1
		<i>N_a</i>	4	5	9	3	1	2	4	1	1	3.3
		<i>H_o</i>	0.167	0.615	0.500	0.091	0.000	0.000	0.125	0.000	0.000	0.166
		<i>H_e</i>	0.642	0.751	0.805	0.376	0.000	0.375	0.492	0.000	0.000	0.382
Harlan, NE 40.0426° -99.2500°	2002	<i>N</i>	18	18	13	19	15	16	11	11	16	15.2
		<i>N_a</i>	2	3	3	1	2	2	2	1	1	1.9
		<i>H_o</i>	0.556	0.556	0.692	0.000	0.000	0.125	0.455	0.000	0.000	0.265
		<i>H_e</i>	0.401	0.579	0.565	0.000	0.444	0.430	0.351	0.000	0.000	0.308
Bull Shoals, MO 36.4037° -92.5998°	1993	<i>N</i>	7	12	10	12	8	12	9	9	9	9.8
		<i>N_a</i>	2	3	3	1	2	2	3	1	1	2.0
		<i>H_o</i>	0.143	0.333	0.700	0.000	0.000	0.083	0.444	0.000	0.000	0.189
		<i>H_e</i>	0.133	0.455	0.660	0.000	0.219	0.080	0.494	0.000	0.000	0.227
Millwood, AR 33.7452° -93.9790°	1995	<i>N</i>	15	15	15	14	13	15	15	15	12	14.3
		<i>N_a</i>	2	3	3	1	1	2	3	1	1	1.9
		<i>H_o</i>	0.067	0.667	0.200	0.000	0.000	0.533	0.667	0.000	0.000	0.237
		<i>H_e</i>	0.064	0.518	0.331	0.000	0.000	0.480	0.584	0.000	0.000	0.220
Texoma, OK 33.8482° -96.7189°	1991	<i>N</i>	10	4	9	6	10	4	15	10	15	9.2
		<i>N_a</i>	2	2	3	1	1	2	3	1	1	1.8
		<i>H_o</i>	0.500	0.250	0.333	0.000	0.000	0.250	0.467	0.000	0.000	0.200
		<i>H_e</i>	0.455	0.469	0.290	0.000	0.000	0.219	0.611	0.000	0.000	0.227
Ray-Roberts, TX 33.4042° -97.1051°	1995	<i>N</i>	16	16	17	14	11	13	13	13	14	14.1
		<i>N_a</i>	2	3	3	1	2	2	3	1	1	2.0
		<i>H_o</i>	0.625	0.188	0.765	0.000	0.091	0.154	1.000	0.000	0.000	0.314
		<i>H_e</i>	0.469	0.389	0.493	0.000	0.087	0.260	0.595	0.000	0.000	0.255
Atchafalaya, LA 30.2510° -91.7510°	1994	<i>N</i>	5	5	5	6	6	6	6	5	5	5.4
		<i>N_a</i>	2	2	3	1	1	1	2	2	2	1.8
		<i>H_o</i>	0.800	0.600	0.800	0.000	0.000	0.000	0.000	0.400	0.400	0.333
		<i>H_e</i>	0.480	0.420	0.560	0.000	0.000	0.000	0.444	0.320	0.320	0.283
Victoria, UGA 0.2790°		<i>N</i>	14	13	18	20	16	20	20	19	19	17.7
		<i>N_a</i>	1	2	2	2	3	2	2	1	1	1.8
		<i>H_o</i>	0.000	0.000						0.000	0.000	0.329
		<i>H_e</i>	0.000	0.473	0.500	0.049	0.119	0.500	0.499	0.000	0.000	0.238
Wiang Kao, THAI 16.1220° 102.5540°		<i>N</i>	8	8	8	8	8	8	8	8	8	8.0
		<i>N_a</i>	2	1	2	1	1	1	2	1	2	1.4
		<i>H_o</i>	0.250	0.000	0.125	0.000	0.000	0.000	0.500	0.000	0.250	0.125
		<i>H_e</i>	0.375	0.000	0.117	0.000	0.000	0.000	0.469	0.000	0.375	0.148
Phewa, NEP 28.2140° 83.9458°		<i>N</i>	18	19	19	20	20	20	8	20	19	18.1
		<i>N_a</i>	2	1	8	1	2	2	4	3	2	2.8
		<i>H_o</i>	0.000	0.000	0.947	0.000	0.050			0.450	0.000	0.236
		<i>H_e</i>	0.105	0.000	0.789	0.000	0.049		0.414	0.359	0.100	0.250
Lyell, AUS -33.5151° 150.7840°		<i>N</i>	3	3	4	4	4	4	4	4	4	3.8
		<i>N_a</i>	2	3	2	1	1	2	1	3	1	1.8
		<i>H_o</i>	0.000			0.000	0.000	1.000	0.000	0.750	0.000	0.269
		<i>H_e</i>	0.444		0.375	0.000	0.000	0.500	0.000	0.656	0.000	0.294
Paramatta, AUS -33.7916° 151.8960°		<i>N</i>	12	13	13	14	14	14	12	14	14	13.3
		<i>N_a</i>	3	3	2	1	3	2	1	5	1	2.3
		<i>H_o</i>	0.417			0.000	0.143		0.000	0.214	0.000	0.232
		<i>H_e</i>	0.601	0.536	0.426	0.000	0.135	0.497	0.000	0.370	0.000	0.285

Grey shading indicates native populations. *N* = number of individuals genotyped for each locus, *N_a* = number of alleles, expected (*H_e*) and observed (*H_o*) heterozygosity. Bold numbers for *H_e* indicate a significant deviation from HW equilibrium (*p* < 0.01). Values calculated with Genalex 6.41
 MN Minnesota, GA Georgia, NE Nebraska, MO Missouri, AR Arkansas, OK Oklahoma, TX Texas, LA Louisiana (all North America), UGA Uganda, THAI Thailand, NEP Nepal, AUS Australia

Fig. 1 Map of sample locations and invasion history of populations in North America (insert map kindly provided by Amy Benson, USGS NAS). Grey shading outside America represent native areas



saturation (Seutin et al. 1991), or in 95 % ethanol, or were frozen at -80°C until DNA extraction. All North American populations were sampled in summer and fall 2008 (except for Atchafalaya, LA, sampled in 1995). All native populations were sampled between 1994 and 1996 (with the exception of Thailand, sampled in 2008). These five samples (Atchafalaya/North America, Lake Victoria/Africa, Lake Phewa/Nepal, and two lakes in Australia) collected in the 1990s are identical to the ones used in Havel et al. (2000). DNA quality and microsatellite amplification did not differ between the older and more recent samples. Although *D. lumholtzi* is widely distributed, its populations typically have a short peak in abundance and are thus difficult to sample. Therefore the number of populations was limited and size of the

samples suitable for DNA analysis was small (between 4 and 20 individuals per population, see Table 1).

Information on the year in which *D. lumholtzi* was first recorded at a given location of the invaded North American range (Table 1; Fig. 1) was obtained from the Nonindigenous Aquatic Species resources website of the United States Geological Survey (<http://nas.er.usgs.gov>). Although the date of first record is not equivalent to the unknown date of introduction, it is a good estimator on when populations first became large enough to be detectable. The peculiar morphology of this species relative to native *Daphnia* (Havel and Hebert 1993) makes *D. lumholtzi* easy to spot when it is abundant enough to appear in samples. Date of first record allowed us to estimate population age of a given location at the time of sampling.

Molecular analysis

DNA was extracted by either of two methods. First, we used a modification of the HotShot protocol of Montero-Pau et al. (2008), which is optimized for the extraction of DNA from small zooplankton eggs. Modifications of the HotShot method consisted of doubling up the recommended volume of the two buffers when extracting DNA from adult individuals. Second, we used a modified CTAB method (Hillis et al. 1990), modified by D. J. Taylor (personal communication) for individual daphniids by reducing the volume of the extraction buffers by ~50 % and re-suspending the DNA in ~40 µl of sterile water.

Adult *D. lumholtzi* individuals (total number of individuals = 178) were genotyped at 9 polymorphic microsatellite loci using primers originally developed for *Daphnia magna*. These primers were Dma1 (forward 5'-GCAAGCGACTTCCACTTTGG, reverse 5'-GTGGCTAACATACTCGGTCG) and Dma3R (forward 5'-A AAGGAAAGCAACCGCTGC, reverse 5'-AAAAGGAAGGGAATTACCC), (primer sequences unpublished, designed by J. Colbourne, personal communication), S6-199 (Routtu et al. 2010), B064, B008, B096 (Jansen et al. 2011), and Dma11, Dma14, Dma15 (published in Genbank, Accession numbers AF291911, AF291913, EU131363, respectively). PCR reactions were performed separately for each microsatellite locus in 12 µl reactions using a PCR kit (iQTM supermix by BioRad), and 1–2 µl of DNA template. The thermal profile used (MJ Research PTC 200 Thermo Cycler) consisted of a 3-min denaturation step at 94 °C, followed by 45 cycles of denaturation at 94 °C for 30 s, annealing at 46 °C for 30 s, extension at 72 °C for 30 s, and a final extension step at 72 °C for 5 min. PCR products were electrophoretically separated on an ABI 3730 using the internal size standard LIZ600 (Applied Biosystems) and allele sizes were analyzed using the PeakScanner Software v1.0 (Applied Biosystems).

Statistical analysis

Multi-locus genotype data were analysed for eight North American populations and five populations from the native range (Africa, Asia and Australia) and included all 9 microsatellite loci. Using the software GENALEX 6.41, we calculated locus-specific number of alleles (N_A), expected (H_e) and observed (H_o) heterozygosity, and tested for departure from

Hardy–Weinberg equilibrium (HWE). Standardized allelic richness and standardized private allelic richness were estimated using the rarefaction method implemented in the software ADZE (Szpiech et al. 2008). Differences in standardized allelic richness and private allelic richness (9 loci) between native and non-native populations were tested with a Mann–Whitney U test. To test for temporal changes in genetic diversity within the invaded range throughout the invasion process, we performed the non-parametric Spearman–Rank correlation of standardized allelic richness with estimated population age, based on the number of years between the first year recorded, and the year of sampling.

The software FreeNA (Chapuis and Estoup 2007) was used to determine the frequency of null alleles for the 9 loci (Online Material 1) and to compute pairwise F_{st} for each population pair with and without using the ENA correction method described in Chapuis and Estoup (2007) and used for graphical display of population differentiation with a Principal Coordinate Analysis (PCoA). To test for a correlation between genetic and geographic distances, we performed a Mantel test with the online software IBDWS (Jensen et al. 2005) on F_{st} values corrected for null alleles (see above) within the invaded range (30,000 randomizations).

We employed a Bayesian clustering algorithm implemented in the software program STRUCTURE 2.3.3 (Pritchard et al. 2000) for the estimation of the number of population clusters (K) using eight introduced and five native populations. For these analyses, three loci were excluded due to missing data (Dma14, B096, Dma 3R). Using multi-locus genotypes, this program assigns individuals to a number of population clusters defined by the user. We calculated estimates for $K = 1$ to $K = 6$. STRUCTURE results are not strongly affected by the presence of null alleles and thus yield correct assignments of individuals when including loci with null alleles (Carlsson 2008). We applied a model allowing for admixture between populations and with correlated allele frequencies, and used sampling locations as priors (the LOCPRIOR model, Hubisz et al. 2009). This method improves the detection of genetic structure if the signal is weak, but is not biased towards detecting a structure if there is none (Hubisz et al. 2009). For each K , 10 replicate simulations were run each with a burn-in of 25,000 iterations and 100,000 Markov-Chain-Monte-Carlo (MCMC) repetitions after burn-in. To evaluate the most likely

number of K , we used the guidelines provided in Pritchard et al. (2000), and the software STRUCTURE HARVESTER (Earl 2011). This online software program applies a method that estimates the “true” K from the greatest rate of change in the likelihood function (ΔK) of successive K values (Evanno et al. 2005). Results of 10 replicate simulations obtained for each K were permuted with CLUMPP (Jakobsson and Rosenberg 2007) for graphical display with DISTRUCT 1.1 (Rosenberg 2004).

To analyze the population genetic structure within the invaded range, we repeated the STRUCTURE analysis outlined above, but we included only the eight North American populations. To test for the influence of mean annual temperature along a latitudinal gradient and invasion history (i.e., year when species was first recorded in a given location) on the population genetic structure in North America, we partitioned the variation obtained in a redundancy analysis (RDA) using the software program CANOCO v.4.5 (ter Braak and Šmilauer 2002). The membership coefficients of individuals (cluster 1–3 obtained from the STRUCTURE analysis) were used as dependent variables in three separate RDAs. We used the two explanatory variables, mean annual temperature and year-of-first-detection together, and separate with either variable as explanatory variable, and covariate to determine the variation explained by either variable, while controlling for the other variable. Monte Carlo permutation tests (9,999 permutations) were performed for each separate RDA to assess the significance of both the first canonical axis and all canonical axes together to evaluate the significance of the variation explained.

Results

Genetic diversity

All 9 microsatellite loci were polymorphic in *D. lumholtzi*, with five to 17 alleles per locus overall and a range of 1–9 alleles per population (Table 1). Standardized allelic richness did not differ between native and invasive populations (Mann–Whitney U test, $Z = 0.732$, $p = 0.46$), while standardized private allelic richness was higher in native populations ($Z = -2.93$, $p = 0.003$, Fig. 2). Standardized allelic richness was negatively correlated with population age (Fig. 3, $r = -0.831$, $p = 0.01$).

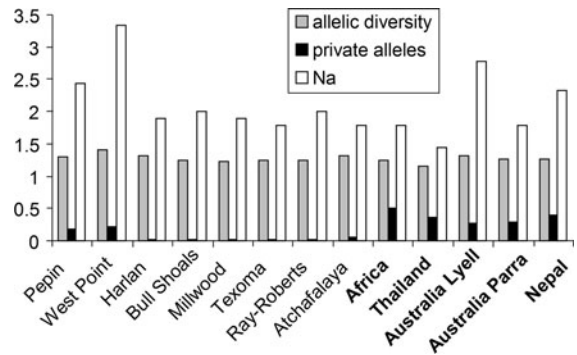


Fig. 2 Allelic patterns at 9 microsatellite loci across populations of the native and invaded range. Standardized allelic richness and standardized private allelic richness were estimated using a rarefaction method with the software ADZE, N_a = total number of alleles detected in each population

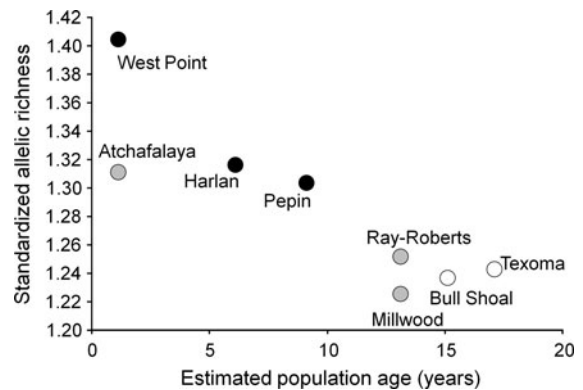


Fig. 3 The relationship between standardized allelic richness with estimated population age (i.e. years since first detection at a given location), estimated by Pearson’s product moment correlation ($r = -0.846$, $p = 0.008$). Datapoints are color-coded by year of detection (white: 1991–1993, grey: 1994–1995, black: 1999–2007)

Genetic distances and genetic structure: linking native and invasive ranges

Pairwise F_{st} values (Table 2; Fig. 4a) were calculated for all populations sampled in the native and non-native regions. Considering the invaded and the native range, the largest genetic distances were found between the two Australian populations and the populations in Arkansas and Oklahoma (ENA corrected $F_{st} > 0.700$). The two smallest genetic distances between the invaded and native range involved the most recently detected population of West Point Lake, USA and the African population of Lake Victoria and the Nepalese population of Lake Phewa

Table 2 Estimates of pairwise F_{st} values for each pair of populations (above diagonal using the ENA correction described in Chapuis and Estoup (2007) between populations of the native and invaded range of *D. lumholtzi*)

	MN	GA	NE	MO	AR	OK	TX	LA	UGA	THAI	AUS1
Pepin, MN		0.134	0.182	0.283	0.322	0.320	0.340	0.355	0.583	0.624	
West Point, GA	0.106		0.152	0.232	0.201	0.235	0.218	0.205		0.467	
Harlan, NE	0.196	0.119		0.209	0.225	0.372	0.260	0.292	0.579	0.514	
Bull Shoals, MO	0.294	0.218	0.214		0.172	0.312	0.127	0.068	0.624	0.553	
Millwood, AR	0.330	0.188	0.221	0.174		0.381	0.192	0.194	0.650	0.607	
Texoma, OK	0.305	0.219	0.391	0.343	0.382		0.307	0.350		0.656	
Ray-Roberts, TX	0.359	0.236	0.271	0.120	0.203	0.309		0.044	0.619	0.539	
Atchafalaya, LA	0.333	0.186	0.270	0.044	0.198	0.346	0.034		0.628	0.541	
Victoria, UGA	0.578		0.589	0.639	0.657	0.665	0.629	0.635		0.677	
Wiang Kao, THAI	0.618		0.511	0.564	0.619	0.663	0.555	0.559			0.581
Phewa, NEP	0.628		0.556	0.596	0.583	0.664	0.587	0.541		0.616	0.675
Lyell, AUS1	0.646		0.635	0.695	0.716	0.699	0.683	0.644		0.762	0.162
Parramatta, AUS2	0.659		0.650	0.680	0.696	0.713	0.670	0.660		0.738	

Grey shading indicates native populations. Abbreviations for geographical locations see Table 1. All values were within the 95 % confidence interval obtained by bootstrapping (FreeNA)

(ENA corrected F_{st} values 0.440 and 0.402 respectively). F_{st} values between more recently invaded populations (Pepin, Harlan, West Point) and native populations were significantly lower (mean 0.555, SE 0.01) than those of older North American populations and native populations (mean 0.626, SE 0.02, Mann–Whitney U test, $Z = -2.72$, $p = 0.007$). In general, pair-wise genetic distances of invaded and native populations were highest when involving Australian populations and smallest when involving African or Asian populations, suggesting Africa and Asia as main source regions. The results of Bayesian clustering obtained from STRUCTURE analysis provide additional evidence similar to that obtained from genetic distance with respect to both potential invasion source regions and relationships between populations in the invaded range (Fig. 5, showing plots for $K = 2$ to $K = 5$). Based on the analysis of an increase of log-likelihood (Evanno et al. 2005), the most probable number of genetic clusters for all 13 studied populations was $K = 2$ (Online Material 2A), clearly differentiating the native from the invaded range, with the exception of the two Australian locations sampled. Following Pritchard et al. (2000), the most probable number of K is located where the distribution of values for the posterior probability of K approaches a plateau, which for this data set is approximately $K = 3$ (Online Material 2B). Increasing the number of genetic

clusters to $K = 3$ distinguished the Australian range from the native populations. A further increase of K yielded the separation of all four native regions at $K = 4$. For all numbers of clusters simulated, the invaded region was clearly dominated by a distinct genetic cluster in most populations of North America with the notable exception of West Point (the most recently detected population). All individuals of West Point shared genetic clusters predominantly with those of the African population of Lake Victoria, and the two Asian populations. To a lesser degree, those clusters were also present in populations located in the Central and Northern USA.

Genetic distances and genetic structure within the invaded range of North America

The smallest genetic distances of the entire study were found within the invaded range of North America (range of corrected F_{st} values 0.044–0.255). Principal Coordinate Analysis (PCoA) of corrected pairwise F_{st} values among populations of the invaded range (Fig. 4b) showed a cluster of the southernmost populations in the United States that included populations in Texas, southern Missouri, Arkansas and Louisiana, while the northern populations in Nebraska (Harlan) and Minnesota (Pepin) and the most recently detected

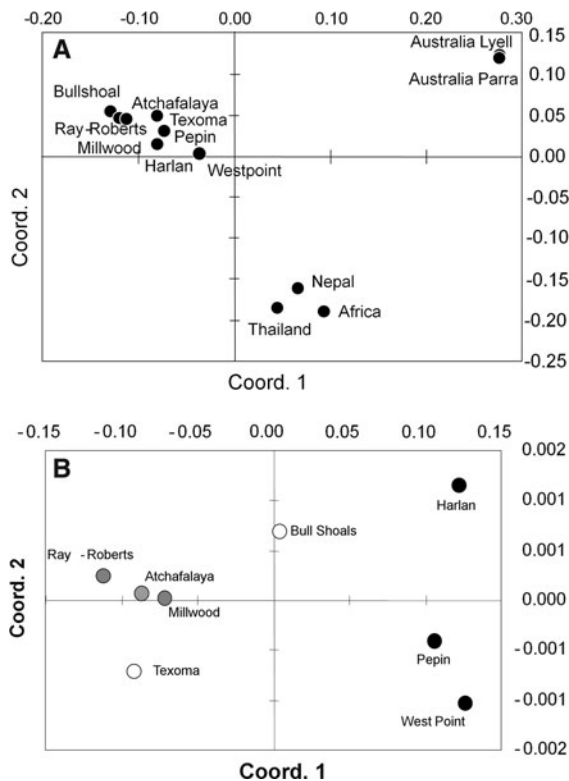


Fig. 4 Principal coordinates analysis (PCoA) of corrected pairwise F_{st} values (ENA correction described in Chapuis and Estoup (2007)) for **a** eight non-native and five native populations of *D. lumholtzi*, and **b** for eight non-native populations. Values of corrected pairwise F_{st} values in Table 2. Color-coding of datapoints as in Fig. 3

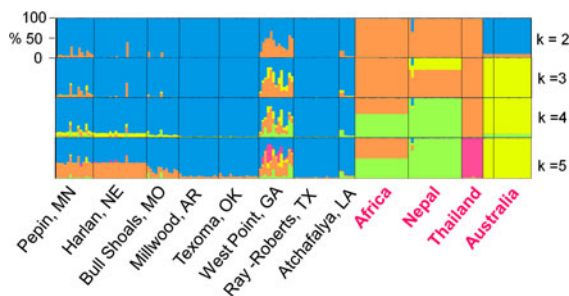


Fig. 5 Population genetic structure of individuals of *D. lumholtzi* from eight North American populations and five populations from its native range in Africa, Asia and Australia. Each vertical bar represents one individual, and colored portions represent assignment probabilities (Q) of individuals to genetic clusters obtained from a series of STRUCTURE analyses with two to five genetic clusters ($K = 2$ to $K = 5$)

population in Georgia (West Point) were more distant from this group. The Mantel test did not show a significant relationship between corrected pairwise

F_{st} values of North American populations and geographical distance ($Z = 7,737.05$, $r = 0.0865$, $p = 0.363$).

In a second STRUCTURE analysis that included all eight North American populations, the true K according to Evanno et al. (2005) was $K = 2$ (Online Material 3A), while following Pritchard et al. (2000), the correct K is higher ($K = 3$, Online Material 3B), indicating substructure that involves more than two clusters (Fig. 6a). The results for $K = 2$ suggest shared ancestry of the most recently detected population of West Point Lake, the northernmost Lake Pepin and the southernmost population of Atchafalaya. Increasing the number of clusters to $K = 3$ showed further structure that appeared to be along a north-south gradient (Fig. 6a, c), with clusters K3 and K1 dominating the two more northern populations (Pepin, Harlan), and cluster K2 dominating the southern populations (Millwood, Texoma, Ray-Roberts and Atchafalaya), with the exception of West Point Lake, which was dominated by cluster K1. For a further study of the effect of mean annual air temperature along a latitudinal gradient and the year of detection on the genetic structure within North America, we employed the more conservative results of $K = 3$ (three genetic cluster membership coefficients of individuals K1, K2 and K3) using RDA (Table 3; Fig. 6b). When we partitioned the variation, 23.5 % ($p = 0.0001$) of the variation in the genetic cluster membership was accounted for by temperature and 34.0 % ($p = 0.0001$) was accounted for by the year in which a population was first detected.

Discussion

We report two novel findings based on multi-locus microsatellite genotypes: first, that *D. lumholtzi* went through multiple introductions to North America from at least two source regions including areas in Asia and Africa; second, that the genetic makeup in the invaded range studied is not uniform, but exhibits a pronounced structure apparently related to both invasion history and a latitudinal temperature gradient. We are aware that the sample size in our study is limited and requires careful interpretation, but nevertheless our results provide a valuable addition to existing knowledge on the distribution and invasion pattern of this exotic zooplankton species.

Fig. 6 Population genetic structure of individuals from eight North American populations and its relation to mean annual temperature and invasion history. The data represented in this graph are obtained from a separate analysis, using only a subset of the entire dataset (see “Methods” for further explanation). **a** Assignment probabilities to genetic clusters for a series of STRUCTURE analyses with two, three and four genetic clusters ($K = 2$, $K = 3$, and $K = 4$). **b** Tri-plot of RDA for three genetic clusters in relation to latitude and invasion history. **c** Map of invaded populations of *D. lumholtzi* in North America. Pie graphs show the assignment probabilities of each population to three genetic clusters ($K = 3$)

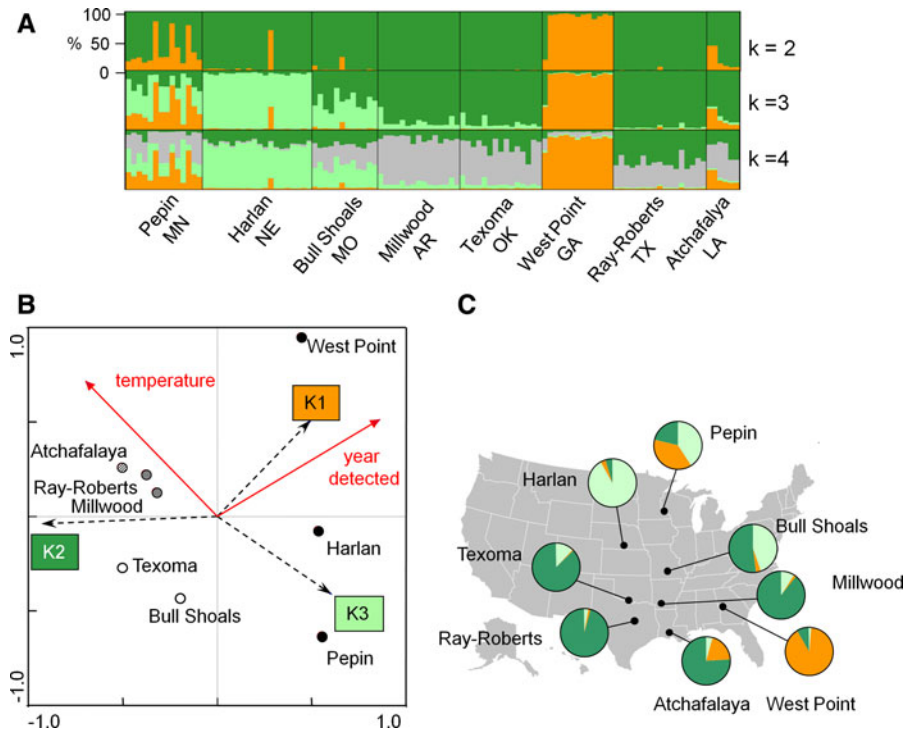


Table 3 Summary of redundancy analysis (RDA) with three genetic clusters identified by STRUCTURE for $K = 3$ as dependent variables, and with mean annual temperature and year-of-detection as explanatory variables

Axes	1	2	3	4	Total variance
Eigenvalues	0.544	0.124	0.252	0.079	1.000
Species-environment correlations	0.930	0.580	0.000	0.000	
Cumulative percentage variance					
of genetic clusters	54.4	66.9	92.1	100.0	
of genetic clusters–environment relation	81.4	100.0	0.0	0.0	
Sum of all eigenvalues					1.000
Sum of all canonical eigenvalues					0.669

The significance of canonical axes was tested with Monte-Carlo tests (9,999 permutations): first canonical axis: F -ratio = 130.085, $p = 0.0001$, all canonical axes: F -ratio = 109.951, $p = 0.0001$

Multiple source introductions

Existing evidence from mtDNA and allozyme data indicated that the most likely introductory source was located in Africa (Havel et al. 2000). The microsatellite data generated in this study confirm this probable source of origin, that introduction events are recent and appear to stem from multiple geographic origins. Pair-wise genetic differences revealed a closer relationship of all North American populations with

populations in Africa and Asia, whereas distances to Australian populations were much larger. Populations from the native range had a significantly higher number of private alleles than in the introduced range, emphasizing the large geographical range and genetic isolation of the native populations. These results are supportive of previous mtDNA data that identified a deep divergence of Australian populations from Asian, African and North American *D. lumholtzi* (Havel et al. 2000). Here, Bayesian clustering of

population microsatellite data resulted in a more complex structure. The four native regions sampled in this study were separated when assuming higher numbers of genetic clusters, while the main “invasive” cluster remained unchanged. Such divergence between native and invaded populations is a common phenomenon of species introductions. In our study, the divergence could suggest either an introduction from unsampled sources, or drift and selection of suitable genotypes after initial invasion and subsequent range expansion in the invaded area, or a combination of both (discussed below). For the identification of more specific introductory sources in future studies, the sampling effort in native regions should be intensified. Nevertheless, despite the modest sample size, we can deduce from our data that North American *D. lumholtzi* does not stem from a single native region, but rather from a mixture of sources including African and Asian populations.

Some populations in the invaded range exhibited shared ancestry with one or more of the native regions. The most recently detected population in West Point Lake (located in the southeastern US) contained the highest percentage of all three native clusters, indicative of its recent invasion from various geographic sources. The three lakes located in the central to northern range in the USA showed similar ancestry with native populations, but were dominated by the North American genetic cluster (the “blue” cluster in Fig. 5). A more recent introduction into West Point is also suggested by the smaller pair-wise genetic distances between populations of the native and invaded range (compared to other North-American populations).

Invasion history in North America: range expansion or secondary introductions?

Although the number of populations sampled for our study was small, the genetic variation detected by microsatellites was sufficient for us to reconstruct the invasion history within North America and its relation to a latitudinal temperature.

Within the invaded range, our data indicate that both invasion history and temperature have played major roles in shaping the genetic structure of these populations. It appears that in *D. lumholtzi*, genetic diversity is similar between native and invaded regions and might only be elevated through admixture

during multiple invasions, as was possibly the case in West Point Lake. According to our results, the most recently detected populations were also the most genetically diverse. A similar relationship of genetic diversity with population age was reported for the invasive lizard *Anolis sagrei*, (Kolbe et al. 2004), for which repeated introductions into new regions within the invaded area, rather than extensive range expansion, were suggested.

Evidence gathered in previous studies points to initial colonization of lakes in the central and southern US (Havel and Hebert 1993), and a subsequent range expansion both north and west across the North American continent (Havel and Shurin 2004). In contrast, the pattern emerging from our study suggests that there were several later introductions. The genetic structure of North American populations, as revealed by Bayesian clustering, appears to be related to temperature as well as to invasion history. In an attempt to disentangle the two factors, we found that in variation partitioning, 24 % of the variation of the individual assignment probability to three genetic clusters in North American populations was explained by temperature, and an additional 34 % by invasion history (the year in which a population was first detected in a location). The significant amount of variation explained by mean annual temperature could reflect adaptation to different climatic zones across the North American continent. Although our data capture the genetic structure at neutral loci, and are as such not an indicator of genetic adaptation to new environments, our study can serve as a basis for designing experimental adaptation studies across climatic zones in North America (e.g. life-history experiments at a range of temperatures).

In its native range, *D. lumholtzi* has a broad distribution in tropical and subtropical regions of Africa, Asia, and Australia (Benzie 2005). In the tropics, the species is perennial or limited to the dry season (Green 1971; Swar and Fernando 1979), and has short peaks in abundance in the subtropics (Timms 1973). Although often abundant during warm periods, the species is also present at lower temperatures in high altitude lakes or other locations such as Lake Phewa in Nepal with minimum temperatures around 16 °C (Green 1995). Such environments could have preadapted *D. lumholtzi* to low winter temperatures even in the southern range of North America, where it can be present temperatures below 6 °C, albeit at low

abundance (Frisch and Weider 2010). In such a scenario, it is possible that the reduction of genetic diversity in older populations resulted from selection of genotypes better adapted to colder climates than those of their original habitats.

Two scenarios might have led to the observed geographical genetic structure: (1) independent invasion of the north and the south and subsequent admixture; or (2) primary introduction to the south, then range expansion and additional recent and repeated introductions from native sources. The first scenario appears to be less likely, since a hypothetical admixture zone in the Central USA should show increased genetic diversity, which was not observed in our study. Instead, our data support the second scenario with the observed decrease in genetic diversity with population age as a result of genetic drift and/or selection, and the presence of native clusters in the more recently invaded populations, which are indicative of ongoing introductions. Populations identified as more recent by observational data in our study all showed an elevated shared ancestry with native populations, indicative of recent invasions, or in the case of West Point Lake repeated invasions, from multiple native sources rather than a range expansion alone. The observation that *D. lumholtzi* can be transported by recreational boats (Havel and Stelzleni-Schwent 2000) has likely contributed to its ability to make long-distance jumps in geographic distribution (Havel and Shurin 2004), including the spread from southern locations northwards. Such a range expansion scenario is supported in the current study by the dominance of a single North American cluster throughout the invaded range, and by the apparent loss of genetic diversity in older populations.

Acknowledgments We are grateful to John Colbourne, Joachim Mergeay and Luc De Meester for sharing microsatellite primer sequences, to Robert Burdis, John Colbourne, Norman Davidson, Meghan Duffy, Wyatt Hoback, Ram Kumar, Chris Luecke, Anke Mueller-Solger, Brian Peterson, La-orsri Sanoamuang, Ron Semyalo, Jim Stoeckel and Colin van Overdijk for providing samples used in this study, and to Amy Benson, United States Geological Survey for providing us with a map of the temporal *D. lumholtzi* distribution in North America. DF was supported by a European Commission Outgoing International Marie Curie postdoctoral fellowship (MOIF-CT-EVOLEXOTIC 40285). We thank two anonymous reviewers for constructive comments on an earlier version of the manuscript.

References

- Ascunce MS, Yang C-C, Oakey J, Calcaterra L, Wu W-J, Shih C-J, Goudet J, Ross KG, Shoemaker D (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science* 331:1066–1068
- Benzie JAH (2005) Cladocera: the genus *Daphnia* (including *Daphniopsis*). Guides to the identification of the microinvertebrates of the continental waters of the world. Backhuys, Leiden, pp 368
- Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *J Hered* 99:616–623
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24:621–631
- Cristescu MEA, Hebert PDN, Witt JDS, MacIsaac HJ, Grigorovich IA (2001) An invasion history for *Cercopagis pengoi* based on mitochondrial gene sequences. *Limnol Oceanogr* 46:224–229
- Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* 17:431–449
- Earl DA (2011) Structure harvester v0.6.1. <http://taylor0.biology.ucla.edu/structureHarvester/>
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* 19:4113–4130
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about biological invasions? *Biol Invasions* 14:245–253
- Frisch D, Weider LJ (2010) Seasonal shifts in genotype frequencies in the invasive cladoceran *Daphnia lumholtzi* in Lake Texoma, USA. *Freshw Biol* 55:1327–1336
- Green J (1971) Associations of Cladocera in the zooplankton of the lake sources of the White Nile. *J Zool Lond* 165:373–414
- Green J (1995) Altitudinal distribution of tropical planktonic Cladocera. *Hydrobiologia* 307:75–84
- Havel JE, Hebert PDN (1993) *Daphnia lumholtzi* in North America—another exotic zooplankton. *Limnol Oceanogr* 38:1823–1827
- Havel JE, Medley KA (2006) Biological invasions across spatial scales: intercontinental, regional, and local dispersal of cladoceran zooplankton. *Biol Invasions* 8:459–473
- Havel JE, Shurin JB (2004) Mechanisms, effects, and scales of dispersal in freshwater zooplankton. *Limnol Oceanogr* 49:1229–1238
- Havel JE, Stelzleni-Schwent J (2000) Zooplankton community structure: the role of dispersal. *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie* 27:3264–3268
- Havel JE, Colbourne JK, Hebert PDN (2000) Reconstructing the history of intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers. *Limnol Oceanogr* 45:1414–1419
- Havel JE, Shurin JB, Jones JR (2002) Estimating dispersal from patterns of spread: spatial and local control of lake invasions. *Ecology* 83:3306–3318

- Hillis DM, Moritz C, Mable BK (1990) Molecular systematics. Sinauer & Associates, Sunderland, USA
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9:1322–1332
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- Jansen B, Geldof S, De Meester L, Orsini L (2011) Isolation and characterization of microsatellite markers in the waterflea *Daphnia magna*. *Mol Ecol Resour* 11:418–421
- Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genet* 6:13. v.3.21. <http://ibdws.sdsu.edu/>
- Kelly DW, Muirhead JR, Heath DD, MacIsaac HJ (2006) Contrasting patterns in genetic diversity following multiple invasions of fresh and brackish waters. *Mol Ecol* 15:3641–3653
- Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177–181
- Lachmuth S, Durka W, Schurr FM (2010) The making of a rapid plant invader: genetic diversity and differentiation in the native and invaded range of *Senecio inaequidens*. *Mol Ecol* 19:3952–3967
- MacIsaac HJ, Grigorovich IA, Hoyle JA, Yan ND, Panov VE (1999) Invasion of Lake Ontario by the Ponto-Caspian predatory cladoceran *Cercopagis pengoi*. *Can J Fish Aquat Sci* 56:1–5
- 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. *Limnol Oceanogr Methods* 6:218–222
- Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic-variability in populations. *Evolution* 29:1–10
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Reusch TBH, Bolte S, Sparwel M, Moss AG, Javidpour J (2010) Microsatellites reveal origin and genetic diversity of Eurasian invasions by one of the world's most notorious marine invader, *Mnemiopsis leidyi* (Ctenophora). *Mol Ecol* 19:2690–2699
- Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol Evol* 22:454–464
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Routtu J, Jansen B, Colson I, De Meester L, Ebert D (2010) The first-generation *Daphnia magna* linkage map. *BMC Genomics* 11:508
- Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Can J Zool* 69:82–90
- Simoes NR, Robertson BA, Lansac-Toha FA, Takahashi EM, Bonecker CC, Velho LFM, Joko CY (2009) Exotic species of zooplankton in the Upper Parana River floodplain, *Daphnia lumholtzi* Sars, 1885 (Crustacea: Branchiopoda). *Braz J Biol* 69:551–558
- Sorensen KH, Sterner RW (1992) Extreme cyclomorphosis in *Daphnia lumholtzi*. *Freshw Biol* 28:257–262
- Swar D, Fernando C (1979) Cladocera from Pokhara Valley, Nepal with notes on distribution. *Hydrobiologia* 66:113–128
- Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24:2498–2504
- ter Braak CJF, Šmilauer P (2002) CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca, New York
- Timms BV (1973) A limnological survey of the freshwater coastal lakes of East Gippsland, Victoria. *Aust J Mar Freshw Resour* 24:1–20
- Yan ND, Leung B, Lewis MA, Peacor SD (2011) The spread, establishment and impacts of the spiny water flea, *Bythotrephes longimanus*, in temperate North America: a synopsis of the special issue. *Biol Invasions* 13:2423–2432
- Zanata LH, Espindola ELG, Rocha O, Pereira RHG (2003) First record of *Daphnia lumholtzi* (Sars, 1885), exotic Cladoceran, in Sao Paulo State (Brazil). *Braz J Biol* 63:717–720