MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1411–1422 doi: 10.1111/j.1365-294X.2011.05465.x

Speciation with gene flow and the genetics of habitat transitions

MELANIA E. CRISTESCU,* ANNA CONSTANTIN,* DAN G. BOCK,* CARLA E. CÁCERES† and TERESA J. CREASE‡

*Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON, Canada N9B 3P4, †Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, ‡Department of Integrative Biology, University of Guelph, Guelph, ON, N1G 2W1, Canada

Abstract

Whether speciation can advance to completion in the face of initially high levels of gene flow is a very controversial topic in evolutionary biology. Extensive gene exchange is generally considered to homogenize populations and counteract divergence. Moreover, the role of introgressive hybridization in evolution remains largely unexplored in animals, particularly in freshwater zooplankton in which allopatric speciation is considered to be the norm. Our work investigates the genetic structure of two young ecological species: the pond species, *Daphnia pulex* and the lake species, *Daphnia pulicaria*. Phylogenetic and population genetics analyses were conducted on mitochondrial *NADH dehydrogenase 5* (ND5) gene, the nuclear *Lactate dehydrogenase* (Ldh) gene and 21 nuclear microsatellite markers in 416 individuals from habitats with various degrees of permanence. The strong and consistent phylogenetic discordance between nuclear and mitochondrial markers suggests a complex evolutionary history of multiple independent habitat transition events that involved hybridization and introgression between lake and pond *Daphnia*. On the other hand, the low level of contemporary gene flow between adjacent populations indicates the presence of effective habitat isolating barriers. The *Daphnia* system provides strong evidence for a divergence-with-gene flow speciation model that involves multiple habitat transition events.

Keywords: adaptive divergence, *Daphnia pulex*, *Daphnia pulicaria*, ecological speciation, gene flow, habitat transition

Received 9 September 2011; revision received 7 December 2011; accepted 14 December 2011

Introduction

The geography and ecology of cladoceran speciation

The relative contribution of geography and ecology to the diversification of freshwater biota is little understood. Allopatric speciation has been regarded as the main mechanism of speciation in many freshwater organisms in which species boundaries often coincide with current or past geographic barriers (Frey 1982; Adamowicz *et al.* 2009; Xu *et al.* 2009). The island-like nature of the limnetic habitat itself creates further opportunity for geographic isolation and restriction of gene flow (Bernatchez & Wilson 1998; Colbourne *et al.* 1998; Waters *et al.* 2001). Recently, evidence of ecologically-based speciation has been revealed in a variety of freshwater species and geographic settings (Lu & Bernatchez 1999; Rundle *et al.* 2007; Schluter & Conte 2009). These finding reinvigorated the long-lasting debate on the relative contribution of geography and ecology to diversification in the freshwater realm. For organisms with exceptionally high dispersal abilities and a cyclically parthenogenetic mode of reproduction, such as anomopod cladocerans, continuous range expansion or colonization of novel aquatic habitats (e.g. habitat transitions) can potentially initiate speciation through a combination of founder effect, natural selection (Lynch 1985; De Meester 1996; Pfrender *et al.* 2000), and hybridization (Taylor & Hebert 1993).

Correspondence: Melania E. Cristescu, Fax: +01 519-971-3616; E-mail: mcris@uwindsor.ca

Reproduction by cyclical parthenogenesis (CP) is considered to be conducive to both colonization of new habitats (founder events) and rapid evolution of ecologically relevant traits (Lynch 1985). During the parthenogenetic phase of reproduction, selection acts on the total phenotype of an individual clone, and not on individual traits/loci. This clonal selection spans the entire genome, resulting in a 'build-up of hidden genetic variance' between clones (Lynch 1985). However, during the sexual phase, linkage disequilibrium between loci is eroded, and the genetic variation accumulated during the parthenogenetic phase is released and expressed phenotypically in the population (Lynch & Gabriel 1983; Lynch 1985). This synergistic selection environment of CP is considered to have the potential to accelerate the evolution of ecologically relevant traits and local adaptation. Moreover, strong local adaptation can result in rapid monopolization of resources by the resident genotypes, which can generate a priority effect against subsequent colonists (De Meester 1996; De Meester *et al.* 2002), as well as adaptive divergence in life history traits (Pfrender *et al.* 2000). For example, one such by-product of local adaptation can be changes in the timing of sexual reproduction itself (temporal isolation), a particularly effective mechanism for restricting gene flow (Deng 1997; Jankowski & Straile 2004).

While the evolutionary mechanisms responsible for decreasing gene flow and creating genetic subdivisions are generally well understood, the potential role of interspecific hybridization and introgression in facilitating colonization of new habitats and ecological speciation remains a very contentious issue in speciation research (Mallet 2007). This problem is particularly difficult to track in the aquatic environment, where hybrid zones are less stable (Petrusek *et al.* 2008), hybrids are distributed along large ecological gradients (Pantel *et al.* 2011) and have an assortment of mating systems. One challenge in studying speciation involves pinpointing the reproductive barriers that initiate speciation, and understanding the succession of these barriers through time, leading to the completion of speciation (Coyne & Orr 2004). The speciation process becomes even more complicated when gradual divergence is punctuated by reticulate evolution. Species-rich complexes that involve relatively young lineages with various degrees of reproductive isolation, sharp genetic subdivision between populations and potential for hybridization and introgression provide very good models for studying the evolutionary forces that drive speciation. Such complex systems offer the opportunity to examine a broad range of genetic divergences from the population to the sister species level. By studying the genomic landscape of genetic differentiation across various geographical scales and ecological settings we can make important inferences about the gradual progression and succession of reproductive barriers.

A new model system for speciation research

After more than 200 years of intense ecological and evolutionary research, the *Daphnia pulex* species complex emerges as a model system not only for ecological genomics (Colbourne *et al.* 2011) but also for speciation studies. The complex includes approximately 12 species (distinct mitochondrial lineages) that inhabit a variety of standing freshwater habitats ranging from temperate and arctic lakes and ponds to coastal or sand dune ponds (Colbourne & Hebert 1996; Colbourne *et al.* 1998; Adamowicz *et al.* 2009; Vergilino *et al.* 2011) and are characterized by various degrees of reproductive isolation and significant intraspecific (among populations) genetic subdivision (Crease *et al.* 1997; Pfrender *et al.* 2000). In North America, the two most prevalent species of the complex are *D. pulex* (Leydig), a temperate pond species, and *Daphnia pulicaria* (Forbes), a widely distributed and geographically structured lake species. The two ecologically distinct species are estimated to have diverged relatively recently, less than 2 Mya, based on mitochondrial markers (Colbourne & Hebert 1996; Colbourne *et al.* 1998), or about 82 kya based on nuclear loci (Omilian & Lynch 2009). The two species show marked differences in life history traits; pond *Daphnia* grow faster, have larger body size, shorter life span, and earlier age of first reproduction than lake *Daphnia* (Dudycha & Tessier 1999; Dudycha 2004). Based on these observations, it has been proposed that current barriers to gene flow between lake and pond populations are largely prezygotic, and involve ecologically-based barriers such as habitat (Lynch 1985; Pfrender *et al.* 2000; Heier & Dudycha 2009) and temporal (allochronic) isolation (Deng 1997).

Ecological studies show that *D. pulex* is present in shallow, fishless, temporary ponds for a short period of time in the spring, while *D. pulicaria* can persist in some stratified lakes year-round (Cáceres & Tessier 2003). In ponds, *Daphnia* feed on detritus and phytoplankton, experience mainly invertebrate predation, and are subjected to long periods of anoxia. In addition, their diapausing (ephippial) eggs often experience seasonal drying and/or freezing. In contrast, *Daphnia* in lakes feed on phytoplankton and use diurnal migration to avoid fish predation and competition from other cladoceran species (Wright & Shapiro 1990). Moreover, ephippial egg production is much more variable in permanent habitats where females can overwinter (Cáceres & Tessier 2004).

Sexually reproducing pulex/pulicaria hybrids, capable of backcrossing, can be successfully produced in

laboratory settings (Heier & Dudycha 2009), but are rarely detected in nature. In contrast, obligately parthenogenetic (OP) hybrids are commonly found in nature in intermediate, deforested habitats such as permanent fishless ponds and lakes (Hebert & Crease 1983), or extremely ephemeral ponds. These asexual hybrids can co-occur regionally or locally with the sexual lineages (Hebert & Finston 2001; Heier & Dudycha 2009), and can have a competitive advantage over the parental species due to potential hybrid vigour (Loaring & Hebert 1981) and demographic advantage (Pantel *et al.* 2011). Although OP hybrids are generally considered to be reproductively isolated from their parental species *via* mating system isolation, many OP clones retain the ability to produce viable and fertile males that have the potential to backcross with the parental species (Innes *et al.* 2000) and maintain interspecific gene flow.

Studying current and historical levels of gene flow between and within ecologically dichotomous habitats can provide a better understanding of the evolutionary history of the *D. pulex* complex. We use a combination of phylogenetic and population genetic analyses on pond and lake populations collected from a focal geographic area. Moreover, we employ a suite of mitochondrial and nuclear markers to estimate the extent of gene flow between ecologically divergent habitats as well as the level of genetic subdivision within each habitat type.

Materials and methods

Sample collection and sexuality tests

D. pulex and *D. pulicaria* were collected from a total of 16 habitats, including 7 ponds ($N = 184$) and 9 lakes ($N = 232$) across Michigan, Illinois, and Ontario (Table 1). Single *Daphnia* females were used to establish clonal lines, hereafter referred to as isolates. In order to maximize both geographic coverage and sampling intensity per habitat, we used two approaches. The phylogenetic reconstructions were based on a broader geographic scale (16 habitats) with lower sampling (mean: 26 individuals per habitat), while the population genetic analyses focused on a restricted geographic area (~200 km in diameter) and intensive sampling (mean: 55.8 individuals per habitat) of three shallow (<1 m deep), ephemeral ponds known to dry completely during most summer seasons, and three relatively deep, thermally stratified lakes with similar zooplankton

Table 1 Habitat location, sampling size, mitochondrial and nuclear diversity of *Daphnia pulex* and *D. pulicaria* populations. *N*, number of individuals screened; N_c, number of individuals analysed after removal of clonal duplicates; N_h, number of haplotypes; *h*, haplotype diversity; π, nucleotide diversity; clade membership based on ND5 phylogeny (Fig. 2b); RM, reproductive mode (CP-cyclical parthenogenesis, OP-obligate parthenogenesis); *Ldh*, Lactate dehydrogenase genotype; A_r, allelic richness; H_E, expected heterozygosity

				mtDNA							Nuclear				
Location	ID	Lat	Lng	*N*	N_c	N_h	*h*	π	Clade	RM	*Ldh*	*N*	N_c	A_r	H_E
Ponds															
Disputed*, ON	Dis	42.17	−83.03	54	54	7	0.735	0.002	A	CP	SS	51	51	5.6	0.603
Solomon*, MI	Sol	42.71	−85.38	55	55	14	0.740	0.006	A	CP	SS	50	50	5.4	0.608
St. Michael*, ON	Stm	42.23	−83.07	33	32	3	0.688	0.003	A	CP	SS	34	33	3.6	0.534
Canard, ON	Can	42.12	−82.98	22	–	–	–	–	A, B	CP/OP	SS/SF	–	–	–	–
Gesto, ON	Ges	42.13	−82.88	2	–	–	–	–	A, B	CP/OP	SF	–	–	–	–
West Gull, MI	Wgu	42.41	−85.44	11	–	–	–	–	A, B	OP	SS	–	–	–	–
Grimey, MI	Grm	42.31	−85.36	6	–	–	–	–	A	OP	SS	–	–	–	–
Total				183	141		0.721	0.004				135	134	4.9	0.582
Lakes															
Lawrence*, MI	Law	42.26	−85.21	92	78	8	0.775	0.025	A, B	CP	FF	86	71	2.6	0.410
Three Lakes 2*, MI	Tlk	42.21	−85.26	37	30	5	0.736	0.008	A, B	CP	FF	36	29	2.5	0.406
Warner*, MI	War	42.28	−85.31	63	60	6	0.299	0.005	A, B	CP	FF	48	41	2.5	0.406
Bassett, MI	Bas	42.40	−85.29	10	–	–	–	–	A, B	CP	FF	–	–	–	–
Little Mill, MI	Mil	42.27	−85.15	13	–	–	–	–	A, D	CP	FF	–	–	–	–
Long, IL	Lng	40.14	−87.44	3	–	–	–	–	A	CP	FF	–	–	–	–
Sportsman, IL	Spm	40.14	−87.44	2	–	–	–	–	A, B	CP	FF	–	–	–	–
Clear, IL	Clr	40.14	−87.44	2	–	–	–	–	A	CP	FF	–	–	–	–
Big Gull, ON	Big	44.88	−78.75	10	–	–	–	–	C, D	CP	FF	–	–	–	–
Total				232	168		0.603	0.013				170	141	2.5	0.407

*Populations surveyed intensely for population genetics analyses.

communities. Since *D. pulex* is known to consist of populations that use cyclical parthenogenesis (CP), obligate parthenogenesis (OP), or mixed reproductive strategies (Hebert & Crease 1983), sexuality tests were conducted on all pond isolates to determine their mode of reproduction, using the method of Innes *et al.* (1986). Isolates reproducing by OP were included only in the mitochondrial phylogenetic reconstruction. All pond populations analysed in our population genetic studies were dominated by or contained exclusively CP lineages. Since we were not able to find CP populations of *D. pulex* in close proximity to the *D. pulicaria* populations, the physical distance between the two types of habitats was not well balanced and resulted in lake populations being closer to each other (within ~10 km radius) than either ponds (~100 km) or pond and lakes (~100 km).

Mitochondrial DNA analysis

Genomic DNA was extracted using the CTAB protocol (Doyle & Doyle 1987). The mitochondrial NADH dehydrogenase 5 (ND5) gene was amplified in 415 pond and lake isolates (for detailed methods, see Appendix A). Unique ND5 haplotypes were identified using DNASP v.5.10.01 (Rozas *et al.* 2003). DNASP was also used to calculate genetic diversity indices such as number of haplotypes (N_h), haplotype (h), and nucleotide (π) diversity. To examine the hierarchical population structure for lake and pond populations, we conducted an analysis of molecular variance (AMOVA) in ARLEQUIN v.3.1 (Excoffier *et al.* 2005).

Phylogenetic analyses

Phylogenetic analyses were performed using neighbour-joining (NJ) analysis in MEGA v.5.0 (Tamura *et al.* 2011) and Bayesian inference (BI) in MRBAYES v.3.1.2 (Ronquist & Huelsenbeck 2003). MODELTEST v.3.7 (Posada & Crandall 1998) was used to select the best model of sequence substitution. The NJ analysis was based on nucleotide distances corrected using the Tamura–Nei model (Tamura & Nei 1993), with confidence levels estimated using 1,000 bootstrap replicates. The Bayesian analyses were performed using the best-fit substitution model as determined by MODELTEST (HKY+G). All searches used random starting trees and employed four independent runs. Trees were sampled every 100th generation for 10^6 generations, and the first 25% of trees were discarded as burn-in. The 50% majority rule consensus tree was generated from the remaining trees and the posterior probability of each node was calculated as the percentage of trees recovering a particular node. European *D. pulex* (GenBank accession number DQ235231) and *Daphnia tenebrosa* (HQ434637) were used as outgroups, while panarctic *D. pulex* (HQ434681), western *D. pulicaria* (HQ434678), and eastern *D. pulicaria* (AB512006) were used as reference sequences.

Nuclear Lactate dehydrogenase genotyping

Previous surveys of allozyme variation have shown that lake populations are generally fixed for an electrophoretically 'fast' (F) allele at the *Lactate dehydrogenase (Ldh)* locus (Hebert *et al.* 1989; Crease *et al.* 2011). Pond populations are either fixed for a 'slow' (S) allele, or are SF heterozygotes, which have been reported to reproduce by OP (Innes *et al.* 1986) and are considered to be F1 hybrids of *D. pulex* and *D. pulicaria* (Hebert & Finston 2001). We designed allele-specific primers based on *Ldh* sequences from both species (Crease *et al.* 2011) to determine the *Ldh* genotype of each isolate (Appendix A).

Microsatellite survey and population genetics analyses

A microsatellite-based population genetic survey was performed on a focal geographic area (southern Ontario and south central Michigan) by intensively sampling 3 ephemeral ponds (N = 135 isolates) and 3 lakes (N = 170 isolates; Table 1). We analysed 21 previously mapped microsatellite markers (Table S1 in Supporting information, Appendix A), located on different linkage groups (Cristescu *et al.* 2006). Neutral markers are known to be useful in identifying the signature of ecological barriers to gene flow (greater differentiation between populations in different environments than within similar environments) when divergent selection is strong and migration is intermediate (Thibert-Plante & Hendry 2010).

Repeated multilocus genotypes were detected using GENALEX v.6 (Peakall & Smouse 2006) and removed from the dataset for all subsequent analyses. Microsatellite data was tested for departures from Hardy-Weinberg equilibrium (HWE) using 10,000 permutations in ARLEQUIN. Levels of significance were adjusted by sequential Bonferroni corrections (Rice 1989) and compared with results from the false discovery rate method (Benjamini & Hochberg 1995). ARLEQUIN was also used to calculate the observed (H_O) and expected (H_E) heterozygosities. Genetic diversity indices of allelic richness (A_r) and number of alleles (A), as well as the inbreeding coefficient (F_{IS}), were estimated in FSTAT v.2.9.3.2 (Goudet 2001). The presence of null alleles was tested with MICROCHECKER v.2.2.0 (Van Oosterhout *et al.* 2004).

Linkage disequilibrium (LD) analysis was performed for all pairs of loci (including the ND5 marker) in each

population, using GENEPOP v.4.0.10 (Raymond & Rousset 1995) after removing low frequency alleles (<10%). A probability test using Markov chains (1000 dememorization steps, 100 batches, 1000 iterations per batch) was conducted to determine the likelihood of pairs of loci being in LD. Significance levels were adjusted as above.

The degree of genetic differentiation between populations was assessed by calculating pairwise F_{ST} values with 10,000 permutations in ARLEQUIN. Additionally, genetic relatedness between individuals was inferred by constructing a NJ dendrogram based on Cavalli Sforza-Edwards chord distances (Cavalli-Sforza & Edwards 1967) in POPULATIONS v.1.2.30 (Langella 1999), and by conducting a factorial correspondence analysis (FCA) in GENETIX v.4.05 (Evanno *et al.* 2005). Admixture between pond and lake populations was assessed using a Bayesian clustering analysis implemented in STRUCTURE v.2.3.1 (Pritchard *et al.* 2000). STRUCTURE uses multilocus genotype data to define a set of populations with distinct allele frequencies (hereafter referred to as clusters), irrespective of sampling location. We assessed the likelihood for models with the number of clusters (K) ranging from 1 to 6 (total sites). For each value of K, we carried out 5 independent runs, with 10^5 generations discarded as burn-in followed by an additional 10^6 generations. The optimal number of clusters was estimated using both the method of Pritchard *et al.* (2000) and Evanno *et al.* (2005).

Emigration and immigration rates between pond and lake populations were inferred using a maximum likelihood method implemented in MIGRATE v.3.0.3 (Beerli 2008), which assumes that all interbreeding populations have been sampled. Despite this limitation, we chose to use this approach since it takes into account historical patterns of gene flow. A Brownian motion mutation model was used with 10 short chains (10^4 iterations), 3 long chains (10^6 iterations) and 10^4 iterations discarded as 'burn-in'.

Results

Phylogenetic analyses of the mitochondrial ND5 gene and nuclear Ldh profiles

The 687- bp final ND5 alignment contained 105 variable sites, of which 53 were parsimony-informative. Within the 415 isolates analysed, we identified a total of 51 haplotypes (GenBank accession numbers JN561018 to JN561068). Among these, 29 were unique to pond populations, 21 to lake populations, and one haplotype was shared between ponds and lakes. Consistent with this observation, haplotype diversity was slightly higher for pond (0.721) than for lake (0.603) populations. By contrast, nucleotide diversity was much lower for ponds (0.004) than for lakes (0.013; Table 1). AMOVA analysis (Table S2, Supporting information) revealed that much of the genetic variance (39.95%) was partitioned between ponds and lakes. A slightly smaller amount of variance was found among populations from the same habitat type (29.37%), and within populations (30.68%).

NJ and BI analyses of the unique ND5 haplotypes recovered four well supported phylogroups (Fig. 1b). The closely related clades A and B correspond to the panarctic *D. pulex* group (Colbourne *et al.* 1998; Paland *et al.* 2005) while clades C and D represent the previously recognized western and eastern *D. pulicaria* (Colbourne *et al.* 1998) groups, respectively. Genetic distances between sister groups range from 1.8% (A/B) to 3.3% (C/D). All phylogroups were characterized by mean within-group genetic distances of less than 1.1%. The mitochondrial phylogenies do not support the reciprocal monophyly of pond and lake isolates. While pond isolates group within the *D. pulex* A and B clades, lake isolates were found either in the *D. pulicaria* clades (C and D) or in *D. pulex* clades A and B (Fig. 1b). With one exception, all pulex/pulicaria hybrids (with SF *Ldh* genotypes and OP reproduction) were detected in ponds.

Ldh genotypes

Large scale allozyme surveys of *Daphnia* populations across North America indicate that several loci (such as *Hex* and *Ldh*) are fixed for different alleles in lake and pond populations (Hebert *et al.* 1989). Consistent with this previous work, we found that lake isolates are invariably homozygous for the *Ldh*-F allele, regardless of their mitochondrial haplotype. Moreover, all pond isolates with CP reproduction were homozygous for the *Ldh*-S allele (Fig. 1b). There was no discordance between *Ldh* genotypes determined using PCR and allozymes.

Population genetic analysis of microsatellite loci

Among 305 isolates analysed, we identified 24 repeated multilocus genotypes, which were encountered from 2 to 4 times and were restricted to single lake or pond populations. A total of 192 alleles were detected across the 21 loci assayed, of which 175 (91%; mean 8 alleles/locus) were found in pond populations and 74 (39%; mean 4 alleles/locus) in lake populations (Table S1, Supporting information). The number of alleles unique to ponds (118) was much higher than those for lakes (17). The allelic richness and expected heterozygosity varied from 3.6 to 5.6 (mean 4.9) and 0.534 to 0.608 (mean 0.582), respectively, in pond populations, and

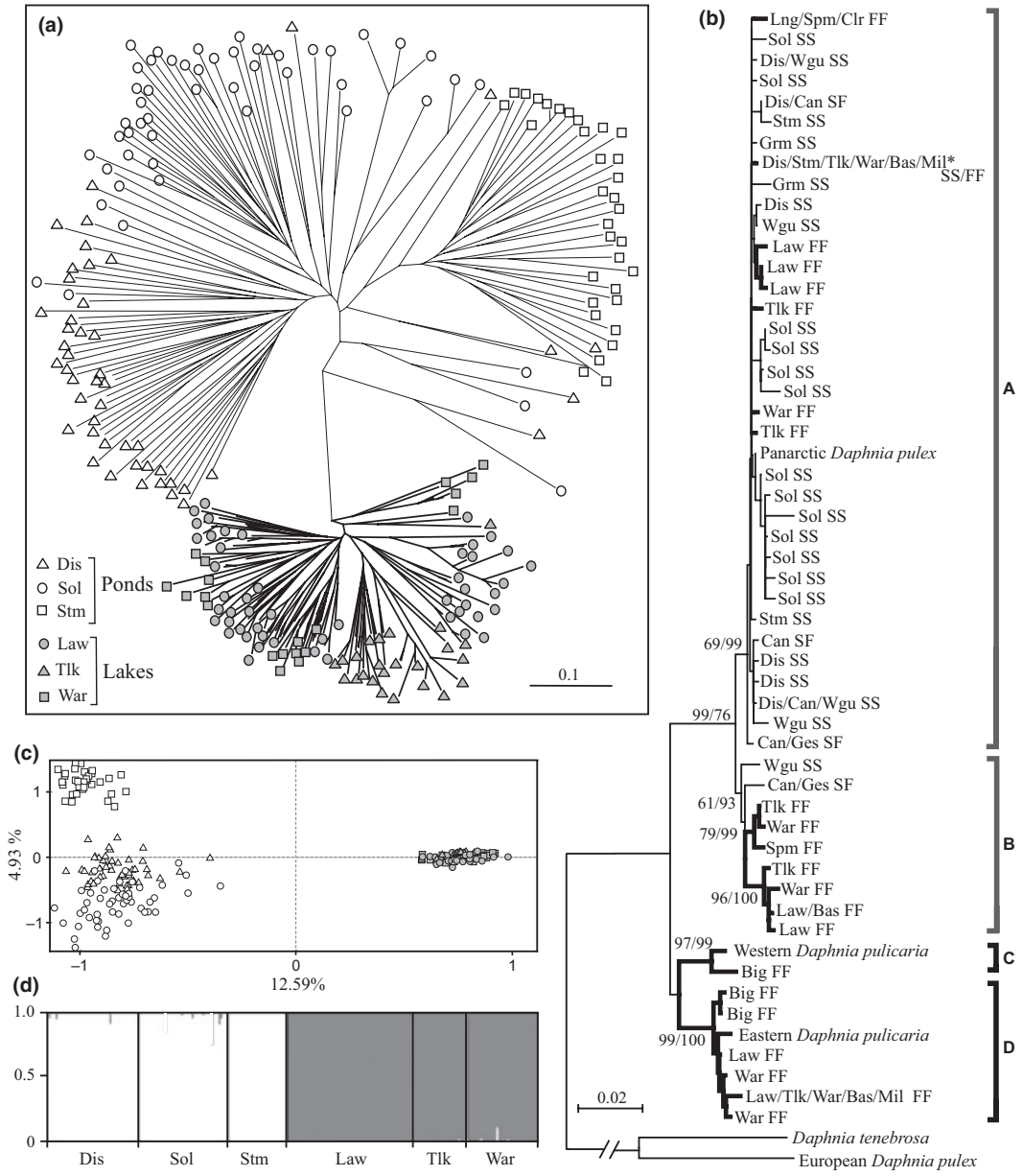

Fig. 1 Analysis of microsatellite and mitochondrial variation in populations of *Daphnia pulex* and *D. pulicaria*. (a) Unrooted NJ dendrogram based on Cavalli–Sforza–Edwards chord distances of genetic variation at 21 microsatellite loci. (b) NJ dendrogram of the mitochondrial ND5 gene. The label of each haplotype indicates its location, followed by *Ldh* genotype (FF, SF or SS). Thick lines denote lake isolates, while thin lines correspond to pond isolates. The single haplotype found in both lakes and ponds is denoted by an asterisk. Numbers beside nodes represent bootstrap support and posterior probabilities, respectively. (c) Factorial correspondence analysis of variation at 21 microsatellite loci for lake and pond isolates. (d) Bayesian STRUCTURE analysis with best support for $K = 2$. The multilocus genotype of each isolate is represented by a vertical line that is partitioned into K shaded segments that represent the isolates' probability of belonging to each of the genetic clusters. Three individuals were removed for this analysis due to relatively high proportion of missing data (23–42%).

from 2.5 to 2.6 (mean 2.5) and 0.406 to 0.410 (mean 0.407), respectively, in lake populations (Table 1). While most loci were in HWE, 24 out of 115 cases showed significant deviations from HWE expectations, including 13 for pond populations, and 11 for lake populations (Table S1, Supporting information). The analysis of LD between pairs of microsatellite loci revealed a much lower number of loci in LD for pond than for lake populations. While ponds had 0.5% of locus pairs in LD, lakes had 2.9–12.1% (mean: 8.8%) of locus pairs in LD.

The analysis of LD between microsatellite loci and the different mitochondrial clades revealed that Lawrence Lake had significantly more non-random associations between microsatellite alleles and mitochondrial haplotypes (43–52% loci in LD) than the other two lakes (0–10% loci in LD; Table S3, Supporting information).

Population differentiation between sampling sites

Microsatellite markers revealed marked genetic differentiation between pond and lake populations with pairwise F_{ST} values ranging from 0.434 to 0.521. By contrast, pairwise F_{ST} estimates between ponds (0.079–0.142) and between lakes (0.085–0.149) were considerably smaller (Table 2). Consistent with this observation, the NJ dendrogram based on the Cavalli Sforza-Edwards chord distances clusters individual genotypes in two distinct groups, with no overlap between lake and pond isolates (Fig. 1a). The FCA analysis also indicated the existence of two main genotype clusters, corresponding to the pond and lake samples (Fig. 1c). All lake genotypes group together irrespective of their mitochondrial haplotype. These results were supported by the STRUCTURE analysis. The most parsimonious model was one with two clusters, corresponding perfectly to the pond and lake groups (Fig. 1d). Furthermore, there was little evidence of contemporary gene flow between ponds and lakes, as both clusters averaged membership coefficients higher than 99%. Analysis of historical gene flow (emigration and immigration rates) averaged over many generations (*Nm*, number of migrants/generation) revealed a relatively high level of gene flow between lakes and ponds (*Nm* range: 0.06–0.76; Table 3).

Table 2 Pairwise F_{ST} estimates between six populations of pond and lake *Daphnia* based on 21 microsatellite loci. All values are significantly different from 0 at $P < 0.05$ and after sequential Bonferroni correction. Site abbreviations are given in Table 1

	Dis	Sol	Stm	Law	Tlk
Sol	0.079	–			
Stm	0.120	0.142	–		
Law	0.481	0.446	0.510	–	
Tlk	0.464	0.434	0.500	0.085	–
War	0.474	0.436	0.521	0.088	0.149

Discussion

Understanding how populations lose evolutionary cohesion, enter divergent evolutionary paths, and eventually complete the speciation process remains enigmatic. The young *D. pulex* species complex offers the opportunity to study the speciation process from the population level to the species level due to its exceptional propensity for both adaptive radiation and hybridization (Colbourne *et al.* 1998; Vergilino *et al.* 2011). This combination of attributes allows the quantification of levels of past and current gene flow among populations or among young species that inhabit standing water bodies with various degrees of permanence and different ecological challenges. Early population studies revealed unexpectedly high levels of genetic structure across very small geographic scales (<100 km) despite the high propensity for dispersal, suggesting that *D. pulex* is perhaps one of the most subdivided species yet studied (Crease *et al.* 1990; Lynch *et al.* 1999) and likely in an active process of geographical and ecological speciation (Pfrender *et al.* 2000; Omilian & Lynch 2009).

D. pulex *as a complex of highly subdivided species*

Our phylogenetic analyses based on the mitochondrial ND5 gene recovered four well supported phylogroups. The two closely related clades A and B had a mixed composition of pond (66.66%) and lake (33.33%) haplotypes (Fig. 1b) and formed a well supported group often recognized as panarctic *D. pulex* (Colbourne *et al.* 1998; Paland *et al.* 2005). The sister clades C and D contained exclusively lake haplotypes (Fig. 1b) known as western and eastern *D. pulicaria* (Colbourne *et al.* 1998).

Table 3 Migration rates (*Nm*, number of migrants per generation) between pond populations of *Daphnia pulex* and lake populations of *D. pulicaria*. Results are averaged over 21 microsatellite loci. Source populations are listed by column, recipient populations listed by row. Site abbreviations are given in Table 1

	Dis	Sol	Stm	Law	Tlk	War
Dis	–	1.6913	1.2790	0.0758	0.5371	0.4423
Sol	1.8408	–	2.6818	0.4812	0.6032	0.7591
Stm	0.7527	0.8021	–	0.0965	0.4910	0.1314
Law	0.3917	0.1799	0.2148	–	1.4034	0.7064
Tlk	0.1643	0.1617	0.1479	1.2166	–	1.0071
War	0.4209	0.1496	0.0552	0.7852	0.7987	–

A parsimony network analysis of the complex Panarctic group further supports the lack of reciprocal monophyly between lake and pond haplotypes (Fig. S1, Supporting information). Furthermore, the network reveals a complex history of habitat transition events, with several distinct groups of lake haplotypes radiating independently from more central haplotypes, suggesting at least three independent habitat transition events. The central haplotype (asterisk in Fig. 1b) of the star-like network was the only haplotype encountered with high frequency in both types of habitats (Fig. S1, Supporting information).

Unlike the mitochondrial marker, the nuclear *Ldh* genotype grouped *Daphnia* isolates consistently by habitat. For example, all lake isolates had the typical (FF) genotype, regardless of their mitochondrial profile. Moreover, our microsatellite data strongly support the reciprocal monophyly of pond and lake *Daphnia* (Fig. 1). These findings are consistent with previous work based on nuclear protein sequences (Omilian & Lynch 2009) and allozymes (Hebert *et al.* 1989; Crease *et al.* 2011) in revealing a strong genetic cohesion across the nuclear genome (coding and non-coding) within each habitat type and significant divergence between ponds and lakes. Most importantly, the discordant phylogenetic signal between the mitochondrial and nuclear genomes (Fig. 1) suggests a history of hybridization and introgression.

Population genetic analysis based on microsatellite data further emphasized the taxonomic complexity of *D. pulex* (sensu lato). While pairwise F_{ST} estimates between populations within each habitat type were moderate (F_{ST} smaller than 0.14; Table 2), marked genetic differentiation was found between pond and lake populations (F_{ST} higher than 0.43). This finding suggests that although gene flow at a local scale was high enough to genetically homogenize populations within each environment, the contemporaneous level of gene flow between ecologically divergent populations is relatively minor. The somewhat unbalanced geographic distribution of lakes and ponds is unlikely to be responsible for the strong genetic structure observed between lakes and ponds. Past allozyme and molecular studies (e.g. Hebert *et al.* 1989; Omilian & Lynch 2009) also revealed a strong association of several nuclear markers with habitat type, irrespective of the geography. Indeed, the pond closest to the lake populations (Sol) was no more similar to the lake populations than were the other two distant pond populations. Our results are consistent with previous findings suggesting that *D. pulex* and *D. pulicaria* are indeed 'good species' and that gene flow is significantly lower between populations in different environments than among populations within each habitat type.

Populations in the two divergent habitats also showed marked differences in several important evolutionary parameters. For example, lake populations were characterized by lower genetic diversity (e.g. haplotype diversity; allelic richness; number of unique alleles; Table 1) and a significantly larger proportion of loci in LD (Table S3, Supporting information). The frequent agreement of genotype frequencies with HWE expectations strongly suggests that the populations analysed are CP. While the majority of microsatellite loci (more than 90%) were in HWE in four of the populations, only about 62% of loci were in HWE in the other two (Solomon Pond and Lawrence Lake). These deviations were characterized by homozygote excess and could be explained by non-random mating (e.g. assortative mating or inbreeding) during the sexual phase, or prolonged periods of clonal selection and/or drift during the parthenogenetic phase of reproduction. The ephemeral ponds sampled in this study have a short season (less than two months) allowing for less than seven parthenogenetic generations per season before the *Daphnia* switch to sexual reproduction and dormancy. Past studies indicate that a short asexual phase, high investment in dormancy and large effective population size are effective in maintaining pond populations close to HWE (Crease *et al.* 1990; Hebert & Finston 2001). We suggest that the moderate deviation from HWE in Solomon Pond, the only pond sampled at the end of the wet season could be attributed to clonal selection. This population showed a slightly higher level of LD between pairs of microsatellite loci (5.2% based on less conservative FDR) than the other pond populations (Table S3, Supporting information). High level of LD is expected after prolonged clonal selection and/or drift.

It is generally assumed that periods of asexual reproduction are significantly longer in stratified lakes than in ephemeral ponds (Morgan *et al.* 2001; Cáceres & Tessier 2004). Detailed studies on the phenology and prevalence of dormancy revealed high variability in the degree of sexual investment among lake populations (Cáceres & Tessier 2004). For example, Cáceres & Tessier (2004) found a lower level of sexual investment in the Lawrence Lake population than either Warner Lake or Three Lakes 2. Prolonged periods of asexual reproduction offer the opportunity for clonal selection and/or drift and can explain, at least partially, the deviation from HWE and the higher level of LD between pairs of microsatellite loci and between certain microsatellite markers and mitochondrial type observed in this population (Tables S3 and S4, Supporting information). These findings suggest that *D. pulex* and *D. pulicaria* populations have experienced different historical-demographic conditions and likely different selection regimes.

Divergence with gene flow in the D. pulex *complex*

Considerable geographic patterning of molecular variation in both mitochondrial and nuclear genomes has often been detected in *D. pulex* populations (Crease *et al.* 1990; Colbourne *et al.* 1998; Paland *et al.* 2005) with greater differentiation across the mitochondrial genome, as expected because of its haploid, uniparental inheritance (Lynch *et al.* 1999). It is generally accepted that the vicariant events associated with the retreat and advancement of glacial ice-sheets initiated divergence between lineages isolated in different glacial refugia (Colbourne *et al.* 1998). For example, divergence between major mitochondrial lineages in the *D. pulex* complex corresponds to about 2 Mya of divergent evolution. However, the contrasting phylogenetic signal between the nuclear and mitochondrial genomes strongly suggests that divergence between *D. pulex* and *D. pulicaria* did not occur by strict allopatric speciation. Instead, we identified a significant level of historical gene flow. A similarly high level of historical gene flow (high enough to prevent divergence in the absence of strong selection) was inferred by Omilian & Lynch (2009) based on sequence variation at six nuclear loci. These findings are consistent with an isolation-with-gene flow scenario. Detailed population genetic studies across pond populations of north-western Oregon revealed no evidence of substantial interbreeding between lake and pond lineages despite the occurrence of lake-like genotypes in ponds following flooding events (Pfrender *et al.* 2000). The findings of Pfrender *et al.* (2000) suggest that the genetic constellation of the pond and lake species remains distinct due to effective ecological barriers to gene flow.

Although hybridization and gene flow (introgression) between species has generally not been considered a constructive force in animal evolution (Mallet 2007), hybridization has been regarded as a potentially important evolutionary process in driving speciation in the ecologically versatile *Daphnia* genus (Hebert 1985; Taylor & Hebert 1993; Schwenk & Spaak 1995). For example, OP *pulex/pulicaria* hybrids (*Ldh*-SF) often detected in ponds have a *D. pulex* mitochondrial haplotype suggesting not only unidirectional hybridization, but also habitat-dependent hybrid parentage that involves females of the common (resident) species and males of the rare (invasive) species (Taylor & Hebert 1993). While heterozygotes for pond and lake alleles at *Ldh* and microsatellite loci were detected only in ponds and not in lakes, many lake isolates contain a *pulex*-like mitochondrial haplotype indicating that hybridization has occurred independently multiple times followed by introgression of *pulex* mitochondrial haplotypes into lake populations.

We suggest that lake-adapted isolates can periodically invade temporary ponds following floods or habitat disruption events. Pfrender *et al.* (2000) also document the frequent invasion of ponds by lake-like *Daphnia* genotypes in north-western Oregon. The lake to pond invasion can be occasionally followed by unidirectional hybridization that involves females of the resident, *D. pulex* species and males produced by the non-resident, *D. pulicaria* invaders (Fig. 2). This scenario explains the absence of lake mitochondrial haplotypes and segregation of lake-adapted nuclear alleles at very low frequency in ponds. Backcrossing of F1s with the resident pond *Daphnia* would produce a wide constellation of F2 and F3 genotypes carrying lake-adapted alleles that, in

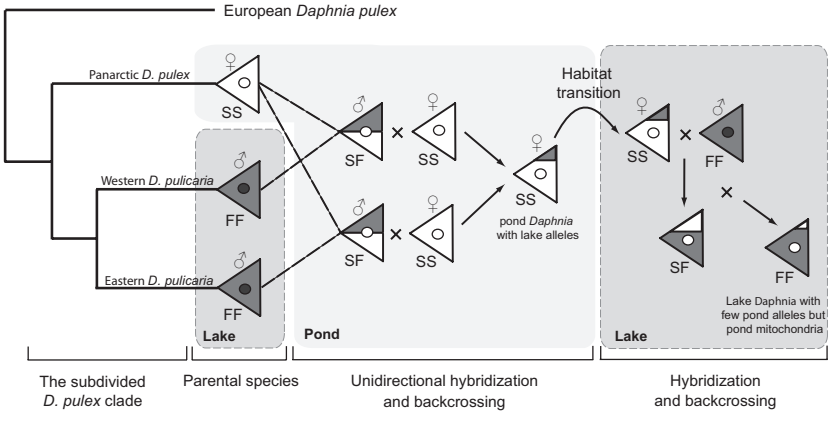

Fig. 2 Proposed model for *Daphnia* speciation and the genetics of habitat transition. White triangles represent *Daphnia pulex* and dark triangles represent *D. pulicaria*. *Lactate dehydrogenase* genotypes (SS, SF, or FF) and mitochondrial profile (white or dark circles) are given for each clade.

rare cases, could provide invasive potential and the ability to withstand habitat transitions. This model explains the rare historical introgression of pond mitochondrial haplotypes into lake populations despite the rarity of hybrids at nuclear loci or the segregation of pond-adapted alleles in permanent, stratified lake populations. We suggest that direct invasion of lakes by pond *Daphnia* followed by successful breeding with lake residents is less likely than this two-step hybridization process, which allows pond *Daphnia* with different life history attributes and naïve to fish predation to 'pre-adapt' and occasionally achieve the transition into lake habitats. A similar model that involves hybridization and backcrossing followed by the transfer of essential alleles across a sharp environmental gradient was proposed by Schluter & Conte (2009) to explain the repeated colonization of freshwater from marine habitats by threespine sticklebacks.

Detailed genomics studies in many species of plants and animals suggest that moderate levels of hybridization and introgression in well-diverged but incomplete species pairs can provide a source of genetic variation that can lead to the creation of beneficial new phenotypes (reviewed by Baack & Rieseberg 2007), which may allow populations to respond to new selection pressures, and thus increases the likelihood of rapid adaptive diversification after invasion of new habitats (Seehausen 2004).

Conclusions

Mitochondrial phylogenetic structure sharply contrasts with that based on nuclear markers (*Ldh* and microsatellites), which consistently separate isolates based on habitat suggesting that the two ecological *Daphnia* species maintain strong intraspecific cohesion and are genetically divergent across their entire nuclear genome. The lack of association between mitochondrial haplotype and habitat suggests that hybridization and introgression of pond *D. pulex* genes into the *D. pulicaria* genome has occurred multiple times. The evidence for hybridization and gene flow revealed by phylogenetic analyses, coupled with moderate to low levels of contemporary gene flow between lake and pond populations suggest a divergence-with-gene flow scenario. The *D. pulex* species complex provides an excellent model for investigating the specific conditions in which hybridization acts as: (i) a sink of genetic divergence, (counteracting local adaptation); (ii) a direct mechanism of speciation generating new OP lineages with various genetic assemblages and broad ecological attributes; (iii) an evolutionary force that enhances ecological divergence through reinforcement or (iv) a mechanism that facilitates habitat transitions through increased ecological tolerance (e.g. invasiveness).

Acknowledgements

We thank D. Innes for providing *Ldh* allozyme confirmation, valuable samples and discussions. We also thank D. Heath, K. Spitze, R. Walter, S. Xu, A. Zhan and the anonymous reviewers for providing helpful comments. Many students helped with sampling and determining the breeding mode of *Daphnia* isolates. Special thanks to C. Perissinotti and A. Waugh. This work was supported by a Natural Sciences and Engineering Research Council (Canada) Discovery Grant and an Early Researcher award from Ontario Ministry of Research and Innovation to MEC.

References

Adamowicz SJ, Petrusek A, Colbourne JK, Hebert PDN, Witt JDS (2009) The scale of divergence: a phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, **50**, 423–436.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, **17**, 513–518.

Beerli P (2008) *MIGRATE version 3.0* (http://popgen.sc.fsu.edu/Migrate-n.html).

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, Université de Montpellier, Montpellier (France).

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Methodological*, **57**, 289–300.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.

Cáceres CE, Tessier AJ (2003) How long to rest: the ecology of optimal dormancy and environmental constraint. *Ecology*, **84**, 1189–1198.

Cáceres CE, Tessier AJ (2004) To sink or swim: variable diapause strategies among *Daphnia* species. *Limnology and Oceanography*, **49**, 1333–1340.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **32**, 550–570.

Colbourne JK, Hebert PDN (1996) The systematic of North American *Daphnia* (Crustacea: Anomopoda): a molecular phylogenetic approach. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **351**, 349–360.

Colbourne JK, Crease TJ, Weider LJ, Hebert PDN, Dufresne F, Hobaek A (1998) Phylogenetics and evolution of the circumarctic species complex (Cladocera: *Daphnia pulex*). *Biological Journal of the Linnean Society*, **65**, 347–365.

Colbourne JK *et al*. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555–556.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, MA, USA.

Crease TJ, Lynch M, Sitze K (1990) Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of *Daphnia pulex*. *Molecular Biology and Evolution*, **7**, 444–458.

Crease TJ, Floyd R, Cristescu ME, Innes D (2011) Evolutionary factors affecting Lactate dehydrogenase A and B variation in the *Daphnia pulex* species complex. *BMC Evolutionary Biology*, **11**, 212.

Cristescu ME, Colbourne JK, Radivojac J, Lynch M (2006) A microsatellite-based genetic linkage map of the water flea *Daphnia pulex*: on the prospect of crustacean genomics. *Genomics*, **88**, 415–430.

De Meester L (1996) Local genetic differentiation and adaptation in freshwater zooplankton populations: patterns and processes. *Ecoscience*, **3**, 385–399.

De Meester L, Gomez A, Okamura B, Schwenk K (2002) The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecologica*, **23**, 121–135.

Deng H-W (1997) Photoperiodic response of sexual reproduction in the *Daphnia pulex* group is reversed in two distinct habitats. *Limnology and Oceanography*, **42**, 609–611.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Dudycha JL (2004) Mortality dynamics of *Daphnia* in contrasting habitats and their role in ecological divergence. *Freshwater Biology*, **49**, 505–514.

Dudycha JL, Tessier AJ (1999) Natural genetic variation of life span, reproduction, and juvenile growth in *Daphnia*. *Evolution*, **53**, 1744–1756.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver.3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Frey DG (1982) Questions concerning cosmopolitanism in Caldocera. *Archiv für Hydrobiologie*, **93**, 484–902.

Goudet J (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. http://www2.unil.ch/popgen/softwares/fstat.htm

Hebert PDN (1985) Interspecific hybridization between cyclic parthenogens. *Evolution*, **39**, 216–220.

Hebert PDN, Crease T (1983) Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. *Heredity*, **51**, 353–369.

Hebert PDN, Finston TL (2001) Macrogeographic patterns of breeding system diversity in the *Daphnia pulex* group from the United States and Mexico. *Heredity*, **87**, 153–161.

Hebert PDN, Beaton MJ, Schwartz SS, Stanton DJ (1989) Polyphyletic origins of asexuality in *Daphnia pulex*. I. Breeding-system variation and levels of clonal diversity. *Evolution*, **43**, 1004–1015.

Heier CR, Dudycha JL (2009) Ecological speciation in a cyclic parthenogen: sexual capability of experimental hybrids between *Daphnia pulex* and *D. pulicaria*. *Limnology and Oceanography*, **54**, 492–502.

Innes DJ, Schwartz SS, Hebert PDN (1986) Genotypic diversity and variation in mode of reproduction among populations in the *Daphnia pulex* group. *Heredity*, **57**, 345–355.

Innes DJ, Fox CJ, Winsor GL (2000) Avoiding the cost of male in obligately asexual *Daphnia pulex* (Leydig). *Proceedings of the Royal Society B, Biological Sciences*, **267**, 991–997.

Jankowski T, Straile D (2004) Allochronic differentiation among *Daphnia* species, hybrids and backcrosses: the importance of sexual reproduction for population dynamics and genetic architecture. *Journal of Evolutionary Biology*, **17**, 312–321.

Langella O (1999) *Populations 1.2.30* (http://bioinformatics.org/~tryphon/populations/).

Loaring JM, Hebert PDN (1981) Ecological differences among clones of *Daphnia pulex* Leydig. *Oecologia*, **51**, 162–168.

Lu G, Bernatchez L (1999) Correlated tropic specialization and genetic divergence in sympatric whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution*, **53**, 1491–1505.

Lynch M (1985) Speciation in the Cladocera. *Verhandlungen des Internationalen Verein Limnologie*, **22**, 3116–3123.

Lynch M, Gabriel W (1983) Phenotypic evolution and parthenogenesis. *American Naturalist*, **122**, 745–764.

Lynch M, Pfrender M, Spitze K *et al.* (1999) The quantitative and molecular genetic architecture of a subdivided species. *Evolution*, **53**, 100–110.

Mallet J (2007) Hybrid Speciation. *Nature*, **446**, 279–283.

Morgan KK, Hicks J, Spitze K *et al.* (2001) Patterns of genetic architecture for life-history traits and molecular markers in a subdivided species. *Evolution*, **55**, 1753–1761.

Omilian AR, Lynch M (2009) Patterns of interspecific DNA variation in the *Daphnia* nuclear genome. *Genetics*, **182**, 325–336.

Paland S, Colbourne JK, Lynch M (2005) Evolutionary history of contagious asexuality in *Daphnia pulex*. *Evolution*, **59**, 800–813.

Pantel JH, Juenger TE, Leibold MA (2011) Environmental gradients structure *Daphnia pulex* x *pulicaria* clonal distribution. *Journal of Evolutionary Biology*, **24**, 723–732.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Petrusek A, Seda J, Machacek J, Ruthova S, Smilauer P (2008) *Daphnia* hybridization along ecological gradients in pelagic environments: the potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363**, 2931–2941.

Pfrender ME, Spitze K, Lehman N (2000) Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Molecular Ecology*, **9**, 1717–1735.

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

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

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

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

Rundle HD, Nagel L, Wenrick Boughman J, Schluter D (2007) Natural selection and parallel speciation in sympatric sticklebacks. *Science*, **287**, 306–308.

Schluter D, Conte G (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the USA*, **106**, 9955–9962.

Schwenk K, Spaak P (1995) Evolutionary and ecological consequences of interspecific hybridization in cladocerans. *Experientia*, **51**, 465–481.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* (in press).

Taylor DJ, Hebert PDN (1993) Habitat-dependent hybrid parentage and differential introgression between neighboringly sympatric *Daphnia* species. *Proceedings of the National Academy of Sciences of the USA*, **90**, 7079–7083.

Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vergilino R, Markova S, Ventura M, Manca M, Dufresne F (2011) Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. *Molecular Ecology*, **20**, 1191–1207.

Waters JM, Craw D, Youngson JH (2001) Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern drainage connections. *Evolution*, **55**, 1844–1851.

Wright D, Shapiro J (1990) Refuge availability: a key to understanding the summer disappearance of *Daphnia*. *Freshwater Biology*, **24**, 43–62.

Xu S, Hebert PDN, Kotov AA, Cristescu ME (2009) The noncosmopolitanism paradigm of freshwater zooplankton: insights from the global phylogeography of the predatory cladoceran *Polyphemus pediculus* (Linnaeus, 1761) (Crustacea, Onychopoda). *Molecular Ecology*, **18**, 5161–5179.

Data accessibility

DNA sequences: Genbank accessions JN561018 to JN561068; Microsatellite data: DRYAD entry doi:10.5061/dryad.f2q28dq4.

Online supplemental material contains: (i) a table that links the individuals sequenced for the population genetics analyses with the GenBank accession numbers; (ii) FASTA-formatted ND5 files for each population; (iii) genetic diversity at 21 microsatellite loci for six *Daphnia* populations; (iv) hierarchical AMOVA analysis based on ND5 data for all populations; (v) Linkage disequilibrium between all pairs of loci in each population; (vi) linkage disequilibrium between microsatellite loci and mitochondrial profiles in lake *Daphnia* and (vii) unrooted statistical parsimony network of mitochondrial ND5 haplotypes of *Daphnia* isolates from the Panarctic clade.

Supporting information

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

Appendix A Detailed molecular methods.

Table S1 Genetic diversity at 21 microsatellite loci for 6 populations of *Daphnia pulex* and *Daphnia pulicaria* with *N*, sample size; *A*, number of alleles; *Ar*, allelic richness (with rarefaction to 27 individuals); *H*O, observed heterozygosity; *H*E, expected heterozygosity; *F*IS, inbreeding coefficient; *P*HW, exact *P*-value for Hardy-Weinberg equilibrium; r, frequency of null alleles. Significant deviations are indicated in bold for sequential Bonferroni corrections (Rice 1989) and italic for the false discovery rate (FDR; Benjamini & Hochberg 1995). Site IDs in Table 1.

Table S2 Hierarchical AMOVA analysis based on ND5 data for *Daphnia pulex* and *Daphnia pulicaria* populations. Comparisons were performed considering pond and lake population as groups. Significant fixation indices at $P < 0.05$ are indicated with asterisks.

Table S3 Linkage disequilibrium between all pairs of loci in each *Daphnia pulex* and *Daphnia pulicaria* population. Significant *P*-values are given in bold for sequential Bonferroni corrections (Rice 1989) and italic for the false discovery rate (FDR; Benjamini & Hochberg 1995).

Table S4 Linkage disequilibrium between microsatellite loci and mitochondrial profiles in lake *Daphnia*. Clade A/B/D indicates an association (LD) between a given microsatellite locus and any mitochondrial profile (clades A, B or D); clade A/B, indicates LD between a microsatellite locus and either clade A or B (figure 2). Significant *P*-values are indicated in bold for sequential Bonferroni corrections (Rice 1989) and italic for the false discovery rate (FDR; Benjamini & Hochberg 1995).

Fig. S1 Unrooted statistical parsimony network of mitochondrial ND5 haplotypes of *Daphnia pulex* and *Daphnia pulicaria* isolates from the Panarctic clade (see Fig. 1). The network was estimated using TCS 1.0 (Clement *et al.* 2000) under the 95% statistical limits of parsimony using the algorithm of Templeton *et al.* (1992). Circles represent sampled haplotypes and the area of each circle is scaled to represent the relative frequency of a particular haplotype. Small dots on the branches represent single nucleotide difference between haplotypes and represent hypothetical (unsampled) haplotypes. The central haplotype of the network was found in both lakes and ponds. Boxes delineate clades that experienced habitat transition events.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.