Molecular Ecology (2008) **17**, 415–430 doi: 10.1111/j.1365-294X.2007.03480.x

Heterogeneity within the native range: population genetic analyses of sympatric invasive and noninvasive clades of the freshwater invading copepod *Eurytemora affinis*

GESCHE WINKLER,*‡ JULIAN J. DODSON* and CAROL EUNMI LEE†
**Québec-Océan, Département de Biologie, Université Laval, Pavillon Vachon, Québec, Canada G1K 7P4, †Center of Rapid Evolution, Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, WI 53706, USA*

Abstract

Invasive species are often composed of highly differentiated populations or sibling species distributed across their native ranges. This study analysed patterns of distribution and the evolutionary and demographic histories of populations within the native range of the copepod species complex *Eurytemora affinis*. Genetic structure was analysed for samples from 17 locations from both the invaded and native ranges in the St Lawrence River drainage basin, using 652 base pairs of the mitochondrial cytochrome oxidase subunit I gene. This study revealed a high degree of heterogeneity in genetic structure and habitat type in the native range, as well as a bias in the sources of invasive populations. Two genetically distinct clades showed a pattern of niche partitioning within the St Lawrence basin. The noninvasive North Atlantic clade primarily occupied the central portion of the St Lawrence Middle Estuary, whereas the invasive Atlantic clade was more prevalent along the margins, in the upstream reaches of the estuary and downstream salt marshes. Habitat partitioning and genetic subdivision was also present within the Atlantic clade. The freshwater populations were genetically more proximate to the Atlantic clade populations in the estuary than to those in the salt marsh, suggesting the estuary as the source of the invasive populations. The freshwater invading populations showed evidence of a modest population bottleneck. Populations from both clades showed genetic signatures of demographic population expansions that preceded the timing of the last glacial maximum, supporting the St Lawrence as a secondary contact zone between the two clades. Additional analyses on physiological and evolutionary properties of populations in the native range, along with analysis of the selection regime within native habitats, might yield insights into the evolutionary potential to invade.

Keywords: biological invasion, COI, demographic history, Great Lakes, source population, St Lawrence River drainage

Received 30 March 2007; revision accepted 26 June 2007

Introduction

Invasive species are often composed of highly differentiated populations or sibling species distributed across their native ranges (Lee 1999, 2000, 2002; Tsutsui *et al.* 2001; Gelembiuk *et al.* 2006; May *et al.* 2006; Ross *et al.* in press). Intriguing patterns are beginning to emerge, as population genetic analyses are beginning to document biases in the sources of invasions from the native range (Lee 1999; Tsutsui *et al.* 2001; Saltonstall 2002; Meusnier *et al.* 2004; Brown & Idris 2005; Chu *et al.* 2006; May *et al.* 2006; Caldera *et al.* submitted). It appears that particular populations (or sibling species) become invasive, whereas others do not. For example, genetically distinct populations (and sibling species) of the zebra mussel *Dreissena polymorpha* span the Black and Caspian Sea region and ancient brackish and freshwater lakes to the south of the seas. However, invasive populations most likely arose from the Black Sea basin and

Correspondence: Carol E. Lee, E-mail: carollee@wisc.edu; Gesche Winkler, E-mail: gesche-winkler@uquar.qc.ca
‡Current address: Institut des Sciences de la Mer (ISMER), Université du Québec à Rimouski, 310 Allée des UrsuLines, Rimouski QC, G5L 3A1, Canada.

not from the ancient lakes (Gelembiuk *et al*. 2006; May *et al*. 2006). Similarly, the Argentine ant, *Linepithema humile*, and the fire ant, *Solenopsis invicta*, exhibit considerable population genetic structure across their native ranges in South America (Tsutsui *et al*. 2001; Ross *et al*. in press), yet the invasive populations arose from restricted locations within floodplain regions of northeastern Argentina (Tsutsui *et al*. 2001; Caldera *et al*. submitted). What factors enable some populations to invade? Might functional differences among these populations affect their ability to invade? Comprehensive population genetic and functional analyses within the native ranges of invasive species might yield insights into the evolution of invasive populations, but such studies are still rare.

This study analysed the evolutionary and demographic histories of populations within the native range of the copepod species complex *Eurytemora affinis*. This study focused on two genetically distinct clades (or sibling species) that overlap in distribution in the St Lawrence River drainage. The copepod *E. affinis* is a dominant species inhabiting estuaries and salt marshes of the Northern Hemisphere, but has invaded freshwater lakes and reservoirs multiple times independently within the past century as a result of human activity (Fig. 1; Lee 1999, 2000). For example, this copepod extended its range into freshwater reservoirs along the St Lawrence River and into the Great Lakes following the opening of the St Lawrence Seaway *c*. 1959 (Willey 1923; Engel 1952; Anderson & Clayton 1959; Faber *et al*. 1966; Patalas 1972; Lee 1999, 2000). Such invasions might have important implications for disease transmission, as *E. affinis* is a major host of many pathogens, including *Vibrio cholerae*, *V. vulnificus*, and *V. parahaemolyticus* (Huq *et al*. 1983; Heidelberg *et al*. 2002; Colwell 2004).

As is the case for many invasive species, *E. affinis* constitutes a cryptic species complex. Rates of molecular evolution, morphological evolution, and patterns of reproductive isolation are discordant, such that species boundaries are difficult to define (Lee 2000; Lee & Frost 2002; see Methods). *E. affinis* is composed of clades that are morphologically indistinguishable (Lee & Frost 2002), yet with large genetic distances and idiosyncratic patterns of reproductive isolation among the clades (Lee 2000).

Within the North American continent, genetically divergent and reproductively isolated clades overlap in distribution on the Atlantic and Pacific coasts, probably reflecting secondary contact following speciation events (Fig. 1; Lee 2000). Within the St Lawrence River system, two reproductively isolated clades coexist in the native saline habitats of *E. affinis* (Lee 1999, 2000). Most notably, only one of the clades (Atlantic clade) has extended its range into freshwater reservoirs and the Great Lakes following the opening of the St Lawrence Seaway, while the other clade (North Atlantic clade) has remained restricted to its native range (Lee 1999, 2000).

The co-occurrence of invasive and noninvasive clades (or sibling species) within the St Lawrence drainage raises intriguing questions regarding range limits, and factors that allow the ranges to shift. Why did one clade invade fresh water, but not the other? A companion study found that populations from the two clades differed in low salinity tolerance and starvation resistance, in a manner that had implications for their ability to invade fresh water (Skelly *et al*. submitted). Under high food conditions, populations from both clades had similar response to low salinity conditions. However, under low food concentrations the population from the noninvasive clade could not tolerate fresh water (Skelly *et al*. submitted). Given the close geographic proximity of the two clades in the native range, what evolutionary mechanisms might have led to differences in physiological tolerance and in their ability to invade?

The goals of this study were to analyse the genetic structure of populations from invasive and noninvasive clades to gain insights into: (i) finer-scale geographic distribution and levels of genetic subdivision within the St Lawrence drainage and (ii) evolutionary and demographic histories of populations within both clades. Information on geographic distributions of the clades could reveal the nature of the environments that the clades occupy in their native range, and the types of selective pressures they might experience. Evolutionary and demographic histories, as revealed by genetic signatures of population bottlenecks and expansions, would yield clues into the nature of their past environments, such as the degree of instability or constancy (Gelembiuk *et al*. 2006). Past selection regimes could have profound effects on physiological tolerances, evolvability, and invasibility of populations.

We analysed population genetic structure for samples from 17 locations using 652 base pairs of the mitochondrial cytochrome oxidase subunit I (COI) gene. We sampled locations from both the native (estuary and salt marsh pools) and invaded (lakes and reservoirs) ranges within the St Lawrence River drainage (Table 1). Mitochondrial haplotype identity of populations within the St Lawrence drainage yielded information on the geographic distributions of the clades and genetic subdivision within each clade. Extracting population genetic parameters from the genetic data provided information on the evolutionary and demographic histories of populations from each clade.

Materials and methods

Population sampling

To analyse the population genetic structure between and within the sympatric clades of *Eurytemora affinis* within the

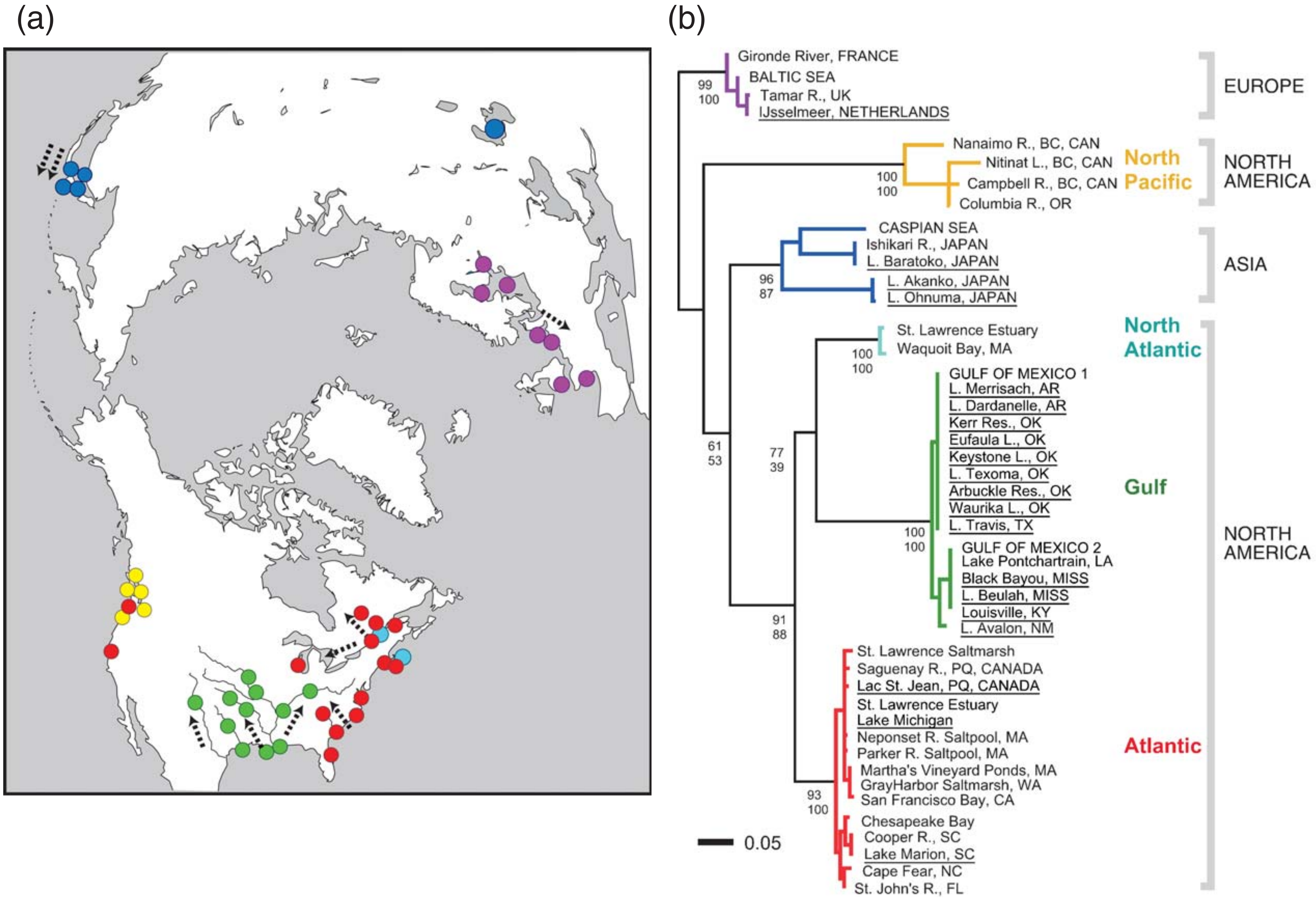


Fig. 1 (a) Geographic pathways of freshwater invasions from saline sources (arrows). (b) Phylogeny of *Eurytemora affinis* populations (using 652 bp COI), where freshwater populations are underlined. Colours represent genetically distinct clades. Figures were adapted from Lee (1999, 2000).

Table 1 Locations sampled within the St Lawrence River System for populations of *Eurytemora affinis*. SLME refers to the St Lawrence Middle Estuary. Sites 1–4 , and 15 are recently invaded freshwater habitats

Sampling locations	Sampling date	Sample size*	Salinity (PSU)†	Tidal influence	Latitude (N)	Longitude (W)	Description of locations
(1) Lake Michigan at Green Bay, WI	1994–1998	2	0	No	44°	88°	From Lee (1999)
(2) Lake Michigan at Racine, WI	5 July 2005	25	0	No	42°43′46″	87°46′44″	Shallow enclosed harbour
(3) Lake Erie at Detroit, MI	2001	8	0	No	42°20′24″	82°55′47″	Enclosed harbour at Windsor
(4) Lac St Pierre, Cornwall, Ontario	29 November 2003	5	0	No	45°01′50″	74°40′03″	Reservoir on St Lawrence R.
(5) Berthier sur Mer, Québec	25 June 2003	26	0–0.5	Yes	46°56′07″	70°44′07″	Shallow bay, southern SLME
(6) Montmagny, Québec	13 June 2004	34	0–0.5	Yes	46°59′26″	70°33′13″	Shallow bay, southern SLME
(7) Cap Brulé, Québec	16 June 2003	32	0.5–20	Yes	47°07′55″	70°41′08″	Deep (24 m) northern channel of SLME
(8) St Jean Port Joli, Québec	13 June 2004	22	0.5–10	Yes	47°12′59″	70°16′22″	Shallow bay, southern SLME
(9) La Pocatière, Québec	26 June 2003	20	10–20	Yes	47°22′08″	70°05′33″	Shallow bay, southern SLME
Baie de L'Isle Verte, Québec							Series of small, shallow,
(10) Pond M1	5 August 2004	2	28	Yes	48°00′20″	69°25′50″	ephemeral tide pools
(11) Pond U2	5 August 2004	28	33	Yes	48°00′14″	69°25′49″	M1-2 = middle intertidal
(12) Pond U1	10 November 2003	5	15	Yes	48°00′14″	69°25′31″	U1-3 = upper intertidal
(13) Pond M2	5 August 2004	1	28	Yes	48°00′28″	69°23′31″	
(14) Pond U3	11 October 2004	36	15	Yes	48°00′13″	69°23′29″	
(15) Lac St Jean, Québec	1994–1998	4	0	No	48°	71°	From Lee (1999); reservoir
(16) Baie de Ha! Ha!, Saguenay River, Québec	1994–1998	4	0.5–20	Yes	48°	70°	From Lee (1999); shallow bay
(17) Rivière du Loup, Québec	26 June 2003	7‡	15–25	Yes	47°46′44″	69°36′19″	Shallow bay, southern SLME

*Number of individuals sequenced per location.

†As recorded at time of sampling.

‡Genetic and morphological analyses identified individuals from this site as *Eurytemora herdmani* rather than *E. affinis*.

St Lawrence drainage, samples from 17 locations were used for this study (Table 1), including 14 locations that were sampled between 2001 and 2005 and three locations from a previous study (sites 1, 15, 16) (Lee 1999). The middle estuary (sites 5–9, 17) and salt marshes (sites 10–14) were assumed to constitute the native range of both the invasive Atlantic and noninvasive North Atlantic clades within the St Lawrence drainage based on sampling studies that preceded the freshwater invasions (Willey 1923). The St Lawrence Middle Estuary is defined as the estuarine transition zone spanning from Ile d'Orléans farther upstream to the downstream locations at Tadoussac (north shore) and Baie de L'Isle Verte (south shore). Recently invaded locations included a reservoir farther upstream near Cornwall (reservoir Lac St Pierre, site 4), a reservoir on the Saguenay River (Lac St Jean, site 15), and the Great Lakes (sites 1–3). Samples were collected with a 63-μm plankton net from the surface layer (1–2 m deep) from shore (sites 4–6, 8), or by boat using a Boston Whaler (sites 7, 9, 17). Five salt marsh ponds farther downstream at Baie de L'Isle Verte (sites 10–14), were sampled by filtering tide pool water collected in a 1-L jar over a 63-μm sieve. These sampling locations varied in salinity and tidal influence (Table 1).

In this study, the clades within the St Lawrence estuary were regarded as sympatric as they were found to co-occur at estuarine and salt marsh locations (sites 7 and 11, respectively). However, we are not certain whether the two clades are truly sympatric in the strict sense of the term. There could be spatial separation due to behavioural or microhabitat differences, which would not be detected using vertical plankton tows. *In situ* sampling would be required to confirm the precise distributions of the two clades within a location.

From hereon, the term 'clade' is used to refer to the Atlantic and North Atlantic clades rather than 'species', because clade designations are easier to define from the phylogenetic relationships (Lee 2000), whereas boundaries for species are unclear. Mating experiments showed that populations from the Atlantic and North Atlantic clades are reproductively isolated and form separate biological species (Lee 2000). However, reproductive isolation was also found between genetically proximate populations within the Atlantic clade (Lee 2000). Therefore, species boundaries are difficult to define. While genetic divergences are very large between the clades (see Results), such high genetic distances are common among morphologically indistinct populations (or sibling species) of copepods (Burton & Lee 1994; Lee & Frost 2002; Goetze 2003; Caudill & Bucklin 2004). The degree to which this pattern is a result of rapid molecular evolution or morphological stasis is unclear.

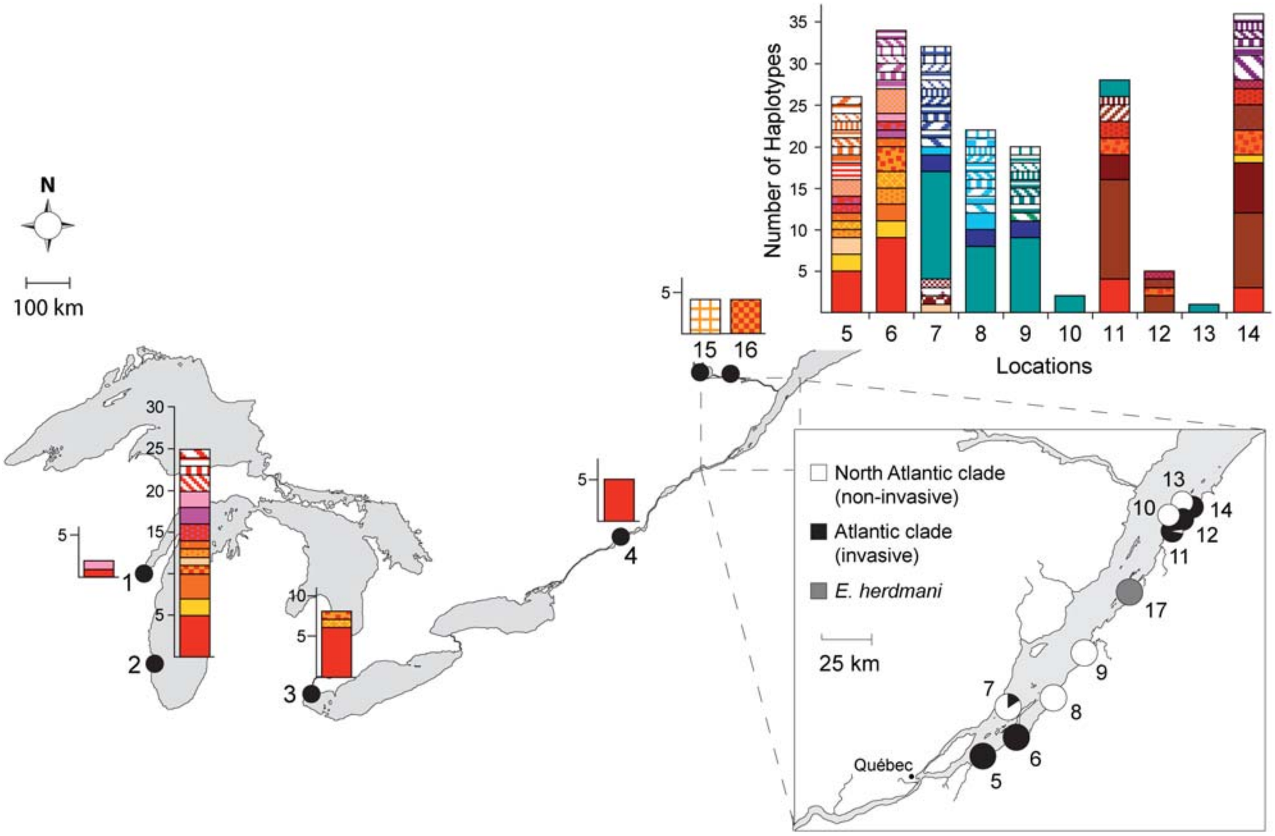


Fig. 2 Geographic distribution and frequency of COI haplotypes for *Eurytemora affinis* from 17 locations within the St Lawrence River drainage basin. Warm (orange, red, purple) colours represent haplotypes from the invasive Atlantic clade, whereas cool (teal, blue) colours represent haplotypes from the noninvasive North Atlantic clade.

DNA sequencing

To determine genetic structure between and within the invasive Atlantic and noninvasive North Atlantic clades in the St Lawrence drainage, DNA sequence data were collected for individual copepods from 17 locations (sample sizes are shown in Table 1; Fig. 2). For each individual, 652 bp of the mitochondrial COI gene was sequenced. DNA was extracted from ethanol-preserved copepods using a cell-lysis buffer with Proteinase-K protocol modified from Hoelzel & Green (1992) (Lee & Frost 2002). Primers COIH 2198 and COIL 1490 (Folmer *et al*. 1994) were used for polymerase chain reaction (PCR) amplification, using a PCR profile taken from Lee (2000). The PCR product was purified by 1.5% agarose gel-purification using a QIAGEN gel-extraction kit. The purified PCR product was cycle-sequenced using ABI Big Dye and DNA sequenced with Applied Biosystems 311 and 3100 sequencers. Complementary strands of most individuals were sequenced to confirm accuracy.

Haplotype diversity and haplotype networks

Estimators of genetic diversity were calculated using DNASP version 4.0 (Rozas *et al*. 2003), including haplotype diversity (Hd, probability that two randomly chosen haplotypes are different in the sample; Nei 1987), nucleotide diversity (π, average number of nucleotide differences per site between two sequences; Nei 1987), and Watterson's θ, based on number of segregating sites (Watterson 1975). Statistical parsimony haplotype networks were constructed to show genetic relatedness among extant haplotypes within the Atlantic (sites 1–7, 11–12, 14–16) and the North Atlantic (sites 7–11, 13) clades using the software package TCS 1.21 (Clement *et al*. 2000).

Analysis of molecular covariance

The hierarchical distribution of genetic covariance was assessed for populations of *E. affinis* (Table 1) by performing an analysis of molecular covariance (AMOVA) using the software package ARLEQUIN version 3.01 (Excoffier *et al*.

2005). The AMOVA was performed at two hierarchical scales: (i) between the Atlantic and North Atlantic clades, and (ii) between estuarine (sites 5–7) vs. salt-marsh populations (sites 11, 12, 14) within the Atlantic clade. In addition, an AMOVA was also performed within the Atlantic clade to test for the more probable saline source of the freshwater-invading populations. Specifically, an AMOVA was performed where the freshwater (invading) populations (sites 1–4) were grouped with either (iii) the salt-marsh populations (sites 11, 12, 14) or (iv) the estuarine populations (sites 5–7). The grouping that combined the invading freshwater populations with the genetically more proximate saline populations (i.e. the more probable source) would be expected to explain a larger share of the covariance (and show larger covariance between the groups). Significance of the observed values of the Φ statistics was tested using a random permutation procedure available in the AMOVA method in ARLEQUIN version 3.01 (Excoffier *et al*. 2005), under the null hypothesis of no population structure.

Pairwise fixation indices (F_{ST}) among populations from all locations were also computed by treating mitochondrial DNA (mtDNA) haplotypes as allelic data at a single locus, and significance was tested by bootstrapping using 1000 permutations (Excoffier *et al*. 2005). The Raymond & Rousset (1995) exact test of population differentiation was used to test for the null hypothesis of identity of haplotype distribution across populations.

Tests of population expansion

Population genetic parameters were determined from the genetic data in order to infer demographic history of the populations. Tests for population growth were performed by coalescent modelling of sequence evolution using the software package LAMARC version 2.1 (Kuhner 2006). Maximum-likelihood estimates were obtained for the joint likelihood surface for θ_1 (= $2N_f\mu$ for mtDNA, where μ = mutation rate, N_f = female effective population size) and population growth rate *g* under a model of exponential growth. The θ_1 computed here is the final θ following population growth. A Metropolis–Hastings Markov chain Monte Carlo algorithm was used, with 10 Markov chains each with 10 000 steps and two final Markov chains each with 200 000 steps. The relationship between θ_0, θ_1, time (*t*) and population growth rate (*g*) is $\theta_0 = \theta_1 e^{-gt}$ (Kuhner 2006). A positive value of *g* indicates population growth, whereas a negative value indicates diminishing population size. Tests were performed for 30 randomly chosen individuals from the Atlantic clade (sites 1–7, 11, 12, 14) and North Atlantic clade (sites 7–11, 13) populations, while all individuals were included in the analyses for freshwater, estuarine, and salt-marsh populations from the Atlantic clade (see Table 3). To test for deviations from a neutral Wright–Fisher model consistent with population expansion, Fu's F_S, Fu and Li's F^*, Fu and Li's D^* and Tajima's D statistic were calculated using DNASP 4.10 (Rozas *et al*. 2003).

The pairwise mismatch distribution was used to estimate the time since demographic expansion, τ (= $2\mu t$, where μ = mutation rate per generation and t = time in generations), as well as initial and final θ, under a model of sudden demographic expansion, using the software package ARLEQUIN version 3.01 (Excoffier *et al*. 2005). A generalized least-squares approach was employed for parameter fitting to the pairwise mismatch distribution. Confidence intervals for the mismatch parameters, τ and initial and final θ, were calculated using 1000 permutations. A goodness-of-fit test was performed to test the validity of the sudden-expansion model, using a parametric bootstrap approach based on the sum of square deviations (SSD) between the observed and expected mismatch distributions (Schneider & Excoffier 1999). The pairwise mismatch analysis was performed for freshwater (sites 1–4), estuarine (sites 5–6), and salt-marsh (sites 11, 12, 14) populations from the Atlantic clade, and for populations from the North Atlantic clade (sites 7–11, 13). We excluded the four Atlantic clade sequences from Cap Brulé (site 7) from the pairwise mismatch analysis because the algorithm for fitting the model to the observed mismatch distribution failed to converge. In order to estimate effective population size and timing of population expansion (in years), a mutation rate of 0.7% per million years was used based on rates of COI sequence divergence in snapping shrimps (Knowlton & Weigt 1998). This rate estimate was used because a calibrated molecular clock does not exist for mtDNA in *E. affinis*. Generation time used was based on six generations per year, observed for *E. affinis* in the St Lawrence Middle Estuary (G. Winkler, unpublished data).

Results

Coexistence of genetically divergent clades within the St Lawrence drainage

Haplotype sequencing of the mitochondrial COI gene from 17 locations within the St Lawrence drainage (Table 1) revealed evidence of habitat partitioning between the Atlantic (invasive) and North Atlantic (noninvasive) clades (Fig. 2). While the two clades showed a pattern of sympatry on a larger geographic scale within the native range, particular locations tended to be heavily dominated by a single clade (Fig. 2).

Within the native range, the noninvasive North Atlantic clade occurred predominantly in the central region of the St Lawrence Middle Estuary (0.5–20 PSU; sites 7–9), including in the deeper northern channel (15–20 m), while the invasive Atlantic clade occurred both upstream and downstream

Table 2 Patterns of mtDNA haplotype variation of *Eurytemora affinis* populations at each location. A = Atlantic clade; NA = North Atlantic clade

Habitat		Population (site no.)	Clade	N	No. of haplotypes	No. of private haplotypes	Haplotype diversity (Hd) ± SE	Nucleotide diversity (π) ± SE
Great Lakes and reservoir (freshwater)		Lake Michigan, Green Bay (1)	A	2	2	0		
		Lake Michigan, Racine (2)	A	25	14	4	0.940 ± 0.024	0.004 ± 0.0005
		Lake Erie (3)	A	8	3	0	0.464 ± 0.200	0.002 ± 0.0005
		Lac St Pierre, reservoir (4)	A	5	1	0		
St Lawrence Middle Estuary	Estuarine (northern channel)	Cap Brulé (7)	A	4	4	3		
			NA	28	15	12	0.791 ± 0.081	0.004 ± 0.001
	Estuarine (south shore)	Berthier sur Mer (5)	A	26	18	9	0.957 ± 0.027	0.005 ± 0.0007
		Montmagny (6)	A	34	18	7	0.918 ± 0.035	0.004 ± 0.0006
		St Jean Port Joli (8)	NA	22	13	10	0.870 ± 0.066	0.004 ± 0.0009
		La Pocatiere (9)	NA	20	11	9	0.805 ± 0.090	0.004 ± 0.001
	Salt Marsh	M1 (10)	NA	2	1	0		
		U2 (11)	A	26	7	2	0.760 ± 0.073	0.005 ± 0.0008
			NA	2	1	0		
		U1 (12)	A	5	4	0		
		M2 (13)	NA	1	1	0		
		U3 (14)	A	36	14	6	0.898 ± 0.030	0.006 ± 0.0005
Saguenay River	Freshwater	Lac St Jean, reservoir (15)	A	4	1	1		
	Estuarine	Baie de Ha! Ha! (16)	A	4	1	0		
Overall			A	179	52		0.930 ± 0.012	0.0054 ± 0.0003
			NA	75	34		0.779 ± 0.051	0.0037 ± 0.0006

of the North Atlantic clade (Fig. 2). Populations within the native range from the Atlantic clade occurred in the oligohaline (low salinity) upstream region of the middle estuary in shallow bays (sites 5, 6), in the salt marshes in the Baie de L'Isle Verte (sites 11, 12, 14), and in the Saguenay River (site 16). The salinity distribution of the Atlantic clade was more extreme than that of the North Atlantic clade, with conductivity fluctuating between 500 µS/cm to approximately 1600 µS/cm in the oligohaline region close to Montmagny (site 6) (Roy 2002), and salinity fluctuating seasonally between 15 PSU and 40 PSU in the salt marsh ponds (M. Ringuette, personal communication). We found only the invasive Atlantic clade in freshwater locations, such as the Great Lakes (Lakes Erie and Michigan; sites 1–3) the reservoir Lac St Pierre at Cornwall (site 4), and Lac St Jean (site 15).

Distribution of these two clades overlapped at Cap Brulé (site 7), in the highly dynamic northern channel, and in the salt marsh of Baie de L'Isle Verte (site 11) (Fig. 2). However, the few individuals from the North Atlantic clade found in the salt marsh ponds (Fig. 2, Table 1, sites 10, 13) could have been washed into the tidal pools from the main estuary, rather than represent reproducing populations. At Rivière du Loup (site 17), neither of these clades was present, but only the congener *Eurytemora herdmani* (Fig. 2).

Patterns of genetic polymorphism within clades

The two clades had a mean sequence divergence at COI of 12.4% (Tamura–Nei model). While haplotype diversity (Hd) was high in both clades, haplotype and nucleotide diversities were higher in the Atlantic clade than in the

Table 3 Estimates of the population genetic parameter θ ($2N_f\mu$), based on number of segregating sites (Watterson 1975) and on coalescent modelling of sequence evolution in LAMARC (Kuhner 2006). Female effective population size (N_f) was calculated from the values of θ. Population growth rate (g) was computed under a model of exponential growth using LAMARC

Populations (sites)	Watterson's $\theta_{\text{per site}}$ ± SD	N_f	LAMARC $\theta_{\text{per site}}$ (95% CF)	N_f	g (95% CF)
Atlantic clade (1–7, 11, 12, 14)	0.01556 ± 0.0039	6.67×10^6	0.0985 (0.050–0.999)	4.22×10^7	1699 (911–3007)
Freshwater (1–4)	0.00721 ± 0.0026	3.09×10^6	0.0681 (0.020–0.764)	2.92×10^7	1734 (1056–4568)
Estuarine (5–7)	0.01362 ± 0.0041	5.84×10^6	0.141 (0.060–0.721)	6.04×10^7	1752 (1118–2911)
Salt marsh (11, 12, 14)	0.00675 ± 0.0023	2.89×10^6	0.0114 (0.0046–0.018)	4.89×10^6	193 (–332–852)
North Atlantic clade (7–11, 13)	0.01098 ± 0.0033	4.71×10^6	0.0203 (0.009–0.071)	8.70×10^6	427 (4.09–2444)

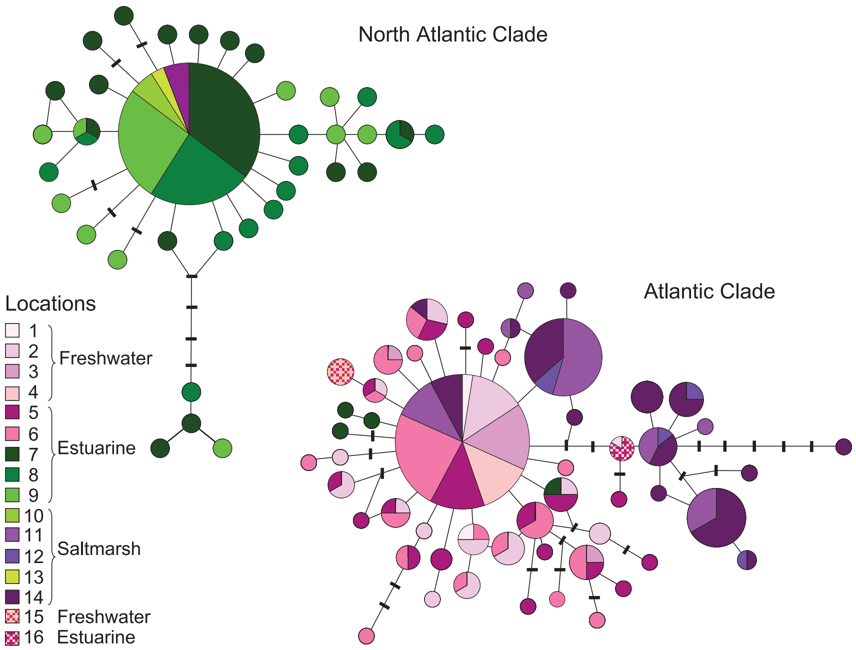


Fig. 3 Parsimony mtDNA haplotype networks for the Atlantic and North Atlantic clades. The areas of the circles are proportional to the frequency of haplotypes. Black bars represent missing haplotypes. Colours indicate the geographic locations of haplotypes. Numbers refer to the same locations shown in Fig. 2 and Table 1.

North Atlantic clade (Table 2). The maximum number of private haplotypes per location was nine for the Atlantic clade and 12 for the North Atlantic clade (Table 2, Fig. 2). Watterson's θ (per site; $2N_f\mu$) was higher in the pooled populations from the Atlantic clade relative to the North Atlantic clade (Table 3), but the values were similar, suggesting similar effective population size.

The haplotype network for the estuarine and freshwater populations from Atlantic clade was star-like, centring around the most frequent haplotype (Fig. 3). In contrast, the haplotype network for the salt-marsh populations from the Atlantic clade was less star-like and more dispersed (Fig. 3). The dominant haplotype in the salt marsh (Fig. 3, sites 11, 12, 14) was only one mutational step away from the most frequent Atlantic clade haplotype. However, this haplotype, and many others in the salt marsh, were not found elsewhere, reflecting the genetic differentiation between the salt marsh and other habitats. The haplotype network for the North Atlantic clade was star-like and centred around the most frequent haplotype, which was found in all the North Atlantic clade populations (Fig. 3; sites 7–11, 13). Nevertheless, there was a satellite cluster 7 mutational steps from the most frequent haplotype, indicating a subgroup within the North Atlantic clade.

Hierarchical patterns of genetic subdivision

We performed an analysis of molecular covariance (AMOVA) to determine the hierarchical distribution of genetic covariance between and within clades (Table 4a), and between and within habitat types within the Atlantic clade (Table 4b). Our AMOVA test revealed that most of the genetic covariance was distributed between the Atlantic and the North Atlantic clades (95.3%, Table 4a). When we compared estuarine vs. salt-marsh populations within the Atlantic clade (sites 5–7 vs. 11, 12, 14), the covariance distributed between the estuarine and salt marsh habitats was substantial (21.9%, Table 4b), although not significant, suggesting genetic subdivision between habitats within the St Lawrence Middle Estuary.

We also performed an AMOVA within the invasive Atlantic clade to identify the saline populations most proximate to the freshwater-invading populations (Tables 4c,d). Our results suggested greater genetic proximity between the freshwater populations (sites 1–4) and those in the estuary (sites 5–7), than between freshwater populations and those in the salt marsh ponds (sites 11, 12, 14). The higher value of Φ_{CT} and the lower value of Φ_{SC}, when the freshwater populations were grouped with the estuarine populations ($\Phi_{CT} = 0.24$, $P = 0.011$; $\Phi_{SC} = 0.02$, $P = 0.099$, Table 4d) than with the salt-marsh populations ($\Phi_{CT} = 0.07$, $P = 0.137$; $\Phi_{SC} = 0.13$, $P < 0.0001$, Table 4c), supported the closer genetic proximity between the freshwater and estuarine populations.

The high pairwise genetic distances between populations (F_{ST}) and the significant P values for the Raymond and Rousset (1995) exact test suggested significant genetic subdivision within the Atlantic clade, but not within the North Atlantic clade (Table 5). The upstream estuarine (sites 5, 6) and Lake Michigan (site 2) populations were genetically distant from the salt-marsh populations (sites 11, 14). The low and nonsignificant pairwise F_{ST} values among populations from Lake Michigan, Berthier, and

Table 4 Analysis of molecular covariance (AMOVA, Excoffier *et al*. 2005) (a) between and within clades and (b) between and within habitat types. Freshwater populations (invasive) were then grouped with salt-marsh populations (c) and with estuarine populations (d) to test for putative source populations. Φ_{CT} is defined as the variance among groups divided by total variance, Φ_{SC} is the variance among populations divided by the variance among and within populations and Φ_{ST} is the variance among groups and among populations divided by total variance (Excoffier *et al*. 2005)

Source of variance	DF	Covariance	% Total	Fixation indices		(*P* value)
(a) Atlantic vs. North Atlantic clades						
Between clades (σ_a^2)	1	33.3	95.3	$\Phi_{CT} = \sigma_a^2/\sigma_T{}^2$	0.95	(< 0.0001)
Among populations within clades (σ_b^2)	16	0.3	0.7	$\Phi_{SC} = \sigma_b^2/(\sigma_b^2 + \sigma_c^2)$	0.16	(< 0.0001)
Within populations (σ_c^2)	236	1.4	3.9	$\Phi_{ST} = (\sigma_a^2 + \sigma_b^2)/\sigma_T{}^2$	0.96	(< 0.0001)
Analyses within the Atlantic clade						
(b) Estuarine (sites 5–7) vs. salt-marsh (sites 11, 12, 14) populations						
Between groups (σ_a^2)	1	0.47	21.9	Φ_{CT}	0.22	(0.091)
Among populations within groups (σ_b^2)	4	0.07	3.3	Φ_{SC}	0.04	(0.052)
Within populations (σ_c^2)	125	1.61	74.9	Φ_{ST}	0.25	(< 0.0001)
(c) Freshwater populations (sites 1–4) grouped with salt-marsh (sites 11, 12, 14) vs. estuarine populations (sites 5–7)						
Between groups (σ_a^2)	1	0.13	7.0	Φ_{CT}	0.07	(0.137)
Among populations within groups (σ_b^2)	8	0.23	12.4	Φ_{SC}	0.13	(< 0.0001)
Within populations (σ_c^2)	161	1.47	80.6	Φ_{ST}	0.19	(< 0.0001)
(d) Freshwater populations (sites 1–4) grouped with estuarine (sites 5–7) vs. salt-marsh (sites 11, 12, 14) populations						
Between groups (σ_a^2)	1	0.48	24.4	Φ_{CT}	0.24	(0.011)
Among populations within groups (σ_b^2)	8	0.03	1.3	Φ_{SC}	0.02	(0.099)
Within populations (σ_c^2)	161	1.47	74.4	Φ_{ST}	0.26	(< 0.0001)

Table 5 Pairwise genetic distances between populations from (a) the invasive Atlantic clade and (b) the noninvasive North Atlantic clade. Estimated pairwise F_{ST} values are indicated in the lower triangular matrix, with asterisks indicating significance level: **P* < 0.05 and ****P* < 0.001. Values in the upper triangular matrix are estimated *P* values of the Raymond & Rousset (1995) exact test of significance (10 000 steps in the Markov chain)

(a) Atlantic clade	Lake Michigan (site 2)	Berthier (site 5)	Montmagny (site 6)	Pond U2 (site 11)	Pond U3 (site 14)
Lake Michigan (site 2)		0.330	0.354	< 0.001	< 0.001
Berthier (site 5)	0.008		0.485	< 0.001	< 0.001
Montmagny (site 6)	0.007	–0.003†		< 0.001	< 0.001
Pond U2 (site 11)	0.192***	0.177***	0.208***		0.354
Pond U3 (site 14)	0.260***	0.252***	0.294***	0.086*	

(b) North Atlantic clade	Cap Brulé (site 7)	St Jean Port Joli (site 8)	La Pocatière (site 9)
Cap Brulé (site 7)		0.718	0.870
St Jean Port Joli (site 8)	–0.014†		0.599
La Pocatière (site 9)	–0.022†	–0.016†	

†Because parameter estimates will often deviate from the true value, a small negative estimate of F_{ST} could be obtained if the true parameter value is zero.

Montmagny (sites 2, 5, 6), suggested a lack of genetic subdivision among them, despite large geographic distances. The low, but significant pairwise F_{ST} value between populations from ponds U2 (site 11) and U3 (site 14), suggests some degree of restricted gene flow at small spatial scales of a couple of hundred metres, or multiple founder effects. However, the Raymond & Rousset (1995) test did not show significant differentiation between these ponds (Table 5a). For the North Atlantic clade estuarine populations (sites 7–9), the low and nonsignificant pairwise genetic distances (F_{ST}) and the high *P* values from the Raymond & Rousset (1995) exact test indicated lack of genetic subdivision (Table 5b), suggesting one panmictic population.

Table 6 Statistical tests of deviation from the standard neutral model, used to test for population growth. Significance values are shown in parentheses. We did not correct for multiple testing

Populations (sites)	Fu's F_S	Fu and Li's F^*	Fu and Li's D^*	Tajima's D
Atlantic clade				
Freshwater (1–4)	–10.22 (<0.0001)	–2.880 (0.008)	–2.631 (0.017)	–2.038 (0.002)
Estuarine (5–7)	–30.27 (<0.0001)	–3.988 (0.002)	–3.946 (0.001)	–2.331 (0.001)
Salt marsh (11, 12, 14)	–2.393 (0.186)	–1.744 (0.070)	–1.997 (0.035)	–0.526 (0.360)
North Atlantic clade (7–11, 13)	–33.97 (<0.0001)	–4.031 (0.001)	–4.087 (<0.0001)	–2.215 (0.001)

Table 7 Population genetic parameters estimated by pairwise haplotype mismatch analysis under a sudden-expansion model. Based on the population genetic parameters, timing of population expansion and female effective population size before (N_{f0}) and after expansion (N_{f1}) were calculated, assuming six generations per year and a mutation rate of 0.7% million/years

	Atlantic clade: freshwater populations (sites 1–4)	Atlantic clade: estuarine populations (sites 5–6)	Atlantic clade: salt-marsh populations (sites 11, 12, 14)	North Atlantic clade: all populations (sites 7–11, 13)
Number of individuals	40	60	67	75
τ	2.34	2.59	6.99	0.64
$\theta_{0\ \text{(per sequence)}}$	0	0	0	1.80
$\theta_{1\ \text{(per sequence)}}$	7.64	104.26	5.86	9.35
$P_{(\text{Sim. SSD} \geq \text{Obs. SSD})}{}^*$	0.83	0.12	0.62	0.93
Timing of population expansion (years BP)	2.57×10^5	2.84×10^5	7.66×10^5	7.01×10^4
CI 5%	7.19×10^4	1.51×10^5	2.59×10^5	0
CI 95%	5.65×10^5	3.97×10^5	1.40×10^6	8.35×10^5
N_{f0}	0	0	0	1.18×10^6
CI 5%	0	0	0	0
CI 95%	9.65×10^5	8.34×10^5	1.73×10^6	3.16×10^6
N_{f1}	5.02×10^6	6.85×10^7	3.85×10^6	6.15×10^6
CI 5%	1.13×10^6	4.19×10^6	1.39×10^6	8.18×10^5
CI 95%	2.36×10^8	9.36×10^8	2.13×10^7	1.23×10^8

*Significance from a goodness-of-fit test of a sudden-expansion model, where a small P value would reject the model.

Population expansion events

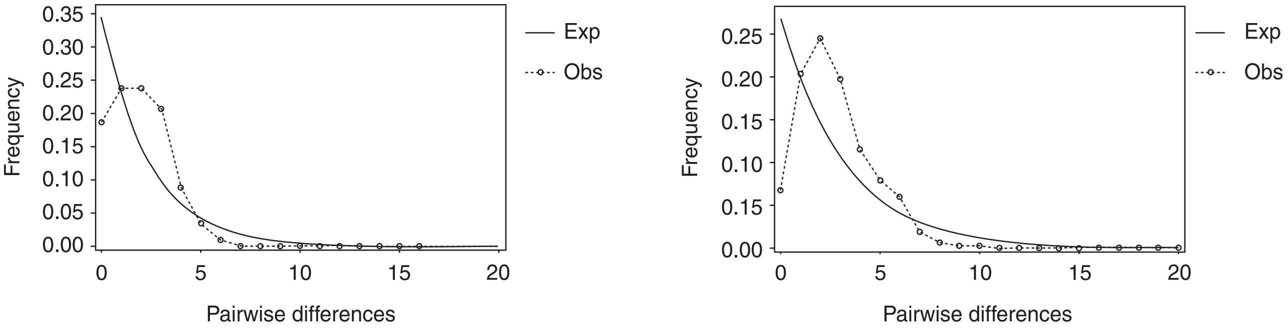
Populations from both clades showed evidence of a history of population expansion events. Freshwater and estuarine populations from the Atlantic clade and populations from the North Atlantic clade showed evidence of population growth based on significant positive values for g (Table 3), significant deviation from a neutral Wright–Fisher model (Table 6), and analyses of pairwise mismatch distributions (Table 7, Fig. 4). Both the freshwater (sites 1–4) and estuarine populations (sites 5–6) from the Atlantic clade showed unimodal mismatch distributions (Fig. 4a, b), consistent with population growth. The pairwise mismatch distribution for the North Atlantic populations (sites 7–11, 13) showed a dominant peak in frequency at two pairwise differences, consistent with population growth (Fig. 4d). A second much smaller peak at seven pairwise differences was consistent with the satellite cluster at seven mutational steps from the most frequent haplotype in the haplotype network (Fig. 3).

For the salt-marsh populations from the Atlantic clade (sites 11, 12, 14), the value for g did not differ significantly from zero (Table 3), and three of the four statistical tests (Fu's F_S, Fu and Li's F^*, Tajima's D) did not deviate significantly from the neutral Wright–Fisher model (Table 6). These salt-marsh populations showed a bimodal distribution (Fig. 4c), which could be consistent with population size constancy, but a model of sudden expansion could not be rejected ($P_{(\text{Sim. SSD} \geq \text{Obs. SSD})} = 0.62$; Table 7).

Mismatch distribution analysis under a sudden-expansion model suggested that timing of a population expansion event occurred earlier in populations from the invasive Atlantic clade than in the noninvasive North Atlantic clade (Table 7). However, confidence intervals were very large. In addition, timing of population expansion appears to have been earlier in the Atlantic clade

(a) Atlantic clade: Freshwater populations

(b) Atlantic clade: Estuarine populations



(c) Atlantic clade: Salt marsh populations

(d) North Atlantic clade

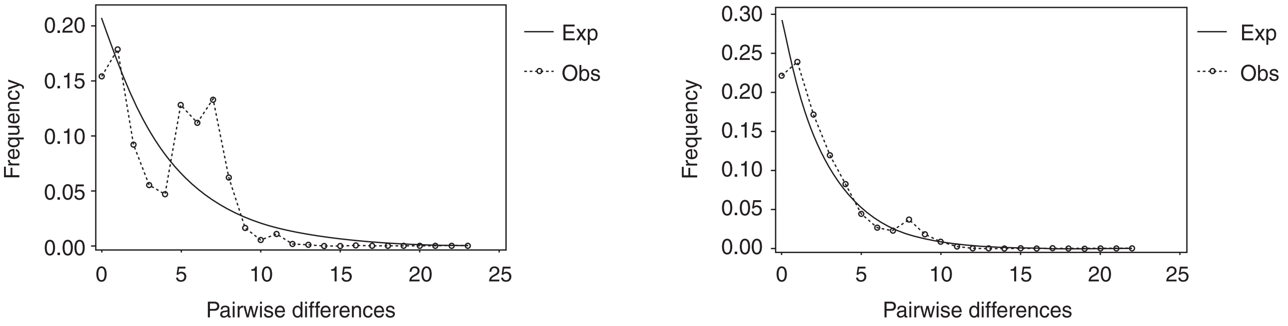


Fig. 4 Pairwise haplotype mismatch distributions for populations from the Atlantic and North Atlantic clades. The solid line represents the expected pairwise mismatch distribution under a Wright-Fisher model, while the dashed lines are the observed frequencies. Graphs show populations from (a) freshwater, Atlantic clade (sites 1–4) (b) estuarine, Atlantic clade (sites 5–6) (c) salt marsh, Atlantic clade (sites 11, 12, 14), and (d) North Atlantic clade (sites 7–11, 13).

salt-marsh populations ($\tau = 6.99$; sites 11, 12, 14) than in the Atlantic clade estuarine populations ($\tau = 2.59$; sites 5–6) (Table 7). Given assumptions of mutation rate and generation time (see Methods), the onset of rapid population growth was ~284 000 years BP for Atlantic clade estuarine populations and ~766 000 BP for the Atlantic clade salt-marsh populations (Table 7; note large confidence intervals). Timing of population expansion for the Atlantic clade freshwater populations (~256 000 BP) was similar to that of the Atlantic clade estuarine populations (Table 7). We estimated a much more recent timing of population expansion for the North Atlantic clade populations of ~70 000 BP (Table 7). Given the possibility of genetic draft (Gillespie 2001), we cannot rule out the possibility that these estimates of demographic expansion actually refer to the timing of the last selective sweep (Bazin *et al.* 2006).

Across all habitats and both clades, estimates of female effective population sizes were moderately large (10^6–10^7; Tables 3 and 7). Estimates of female N_f based on pairwise mismatch analysis (Table 7) and LAMARC (Table 3) tended to be larger than those based on Watterson's θ (Table 3), as the first two methods allow for population growth. The pairwise mismatch analysis in ARLEQUIN assumes a sudden-expansion model, LAMARC assumes continuous population growth up to the present, while Watterson's θ assumes constant population size. The model assumption of sudden population expansion in ARLEQUIN might be the most realistic for this system, given the support for population expansion (Tables 3, 6 and 7) and the geological history of glaciation and recent colonization in this region. Relatively constant estimates for N_f across methods for the Atlantic clade salt-marsh populations (Tables 3 and 7) might result from relative demographic constancy.

Reduction in effective population size (Table 7) and the shift in haplotype frequency (Fig. 2) were consistent with a modest population bottleneck associated with invasion into the Great Lakes. Haplotype distribution between the estuarine (sites 5–7) and freshwater (sites 1–4) populations was significantly distinct (Fisher's exact test, $P = 0.0339$; shared vs. private haplotypes). The core haplotype constituted ~22% in the estuarine populations (i.e. found in 14 out of 64 sampled individuals), but increased to ~43% in the Great Lakes (i.e. found in 17 out of 40 sampled individuals). Also, in Lac St Jean, all four individuals possessed one unique haplotype, consistent with a population bottleneck.

Discussion

Geographic distribution of invasive and noninvasive clades within the St Lawrence drainage

This study revealed a high degree of heterogeneity in genetic structure and habitat type in the native range of an invasive species. The St Lawrence River drainage contained two genetically divergent clades that overlap in distribution, but differ in their potential to invade (Figs 1 and 2). The two clades, the Atlantic (invasive) and North Atlantic (noninvasive), showed evidence of habitat partitioning, where particular locations tended to be heavily dominated by a single clade (Fig. 2). Populations from the Atlantic clade tended to dominate in the more marginal habitats, including upstream oligohaline regions (low salinity, 500–1600 μS/cm) and highly fluctuating salt-marsh ponds (15–40 PSU). In contrast, populations from the North Atlantic clade dominated the central mesohaline portion (0.5–20 PSU) of the middle estuary where food concentrations are sufficient to support a high standing stock of *Eurytemora affinis* in this zone (Vincent *et al*. 1996; Winkler *et al*. 2003).

The geographic partitioning between the two clades was striking, given that strong hydrodynamic forces acting in St Lawrence Middle Estuary would be expected to homogenize the distribution of two clades if the copepods and larvae dispersed as passive particles (Simons *et al*. 2006). Preliminary modelling of passive and active (vertical migrating) particles (Simons *et al*. 2006) showed that particles released in the Northern Channel would disperse throughout the entire brackish zone of the St Lawrence Middle Estuary, including the south shore region (R. Simons and G. Winkler, unpublished data). Behavioural differences between the two clades might be important for achieving separate distributions and maintaining populations within physiologically favourable zones (C. E. Lee, personal observations).

Microhabitat differences due to niche partitioning between the clades in the native range might explain differences in their physiological tolerances and ability to invade fresh water. For example, the reduced freshwater tolerance and lower starvation resistance of the North Atlantic clade might reflect an evolutionary history in a more stable and food-rich habitat (Skelly *et al*. submitted). In contrast, the occurrence of the invasive clade in the marginal habitats, removed from the food-rich central portion of the estuary (Vincent *et al*. 1996; Winkler *et al*. 2003), might select for physiological phenotypes with the capacity to tolerate novel and stressful environments.

Genetic subdivision within the clades

Significant genetic subdivision and habitat partitioning was evident within the invasive Atlantic clade, but not within the noninvasive North Atlantic clade. Within the invasive Atlantic clade, an AMOVA revealed considerable genetic covariance distributed between estuarine (sites 5–7) and salt marsh habitats (sites 11, 12, 14) in the native range (21.9%, Table 4b). Genetic subdivision between estuarine and salt-marsh habitats was also supported by the high and significant pairwise genetic distances between populations (F_{ST}) and low P values from the Raymond and Rousset exact test (1995) (Table 5a). Such genetic subdivision could have arisen from restricted gene flow between these two habitat types. For example, the presence of both the North Atlantic clade and *Eurytemora herdmani* in the central portion of the St Lawrence Middle Estuary (Fig. 2) might have acted as a barrier, possibly through competitive exclusion, between the estuarine and salt-marsh populations of the Atlantic clade.

Results from the Atlantic clade suggest greater demographic constancy over time for the salt-marsh relative to the estuarine populations. The bimodal pairwise mismatch distribution (Fig. 4c), dispersed haplotype network (Fig. 3), low value for g (Table 3), and lack of significant deviation from the standard neutral model for three out of four test statistics (Table 6) support the greater demographic constancy in the salt-marsh populations. Older estimates for timing of population expansion supported longer-term stability, and possibly a more ancient origin, of the salt-marsh populations (Table 7). In contrast, the star-like haplotype network for the estuarine and freshwater populations from Atlantic clade suggested a signal of a more recent population bottleneck and expansion. However, the star-like pattern (Fig. 3) could have also arisen from a selective sweep in the not-too-distant past (Bazin *et al*. 2006).

More extensive sampling of the St Lawrence drainage might reveal additional levels of heterogeneity and population genetic structure within the Atlantic clade. For example, the Atlantic clade population in the deep channel at Cap Brulé (site 7) contained three unique haplotypes out of four individuals, suggesting that populations in the deep channel might be quite diverged from those in other habitats, including those in the shallow regions of the estuary. Significant F_{ST} values between salt-marsh ponds (Table 5a) suggest that additional sampling from the salt marshes might reveal additional genetic structure within and among tidal marsh flats. Additional sampling might also reveal the origins of the Lac St Jean population, which contains one unique haplotype.

In contrast to the genetic subdivision within the Atlantic clade (Table 5a), there was a lack of genetic subdivision within the North Atlantic clade (sites 7–9; Table 5b), consistent with panmixia. The large numbers of private haplotypes found in the North Atlantic clade estuarine populations (Table 2, Fig. 2) were likely an indication of high genetic diversity rather than genetic subdivision. Hydrodynamic conditions of the middle estuary would promote panmixia among estuarine populations, where tides vary between 3 m and 5 m in height, and current speeds may reach 2–3 m/s (along channel) and 0.3–0.5 m/s (cross-channel) (d'Anglejan *et al*. 1981; Simons *et al*. 2006).

Geographic sources of freshwater invasions

Freshwater invasions into the Great Lakes coincided with the opening of the St Lawrence Seaway in 1959 (Engel 1952; Anderson & Clayton 1959; Faber *et al*. 1966; Patalas 1972; Lee 1999). Shipping activity and ballast water transport might have served as the mechanism for the recent invasions. Reduced polymorphism and lower values for θ (Tables 3 and 7) suggested a modest bottleneck following invasions into freshwater habitats (see Results).

Population genetic analyses within the Atlantic clade indicated that the freshwater populations (sites 1–4) were probably derived from the estuary (sites 5–7), rather than from the salt-marsh populations (sites 11, 12, 14). The low pairwise F_{ST} between freshwater (site 2) and estuarine populations (sites 5, 6) (Table 5) supported the lack of genetic differentiation between freshwater and estuarine populations. Likewise, the AMOVA supported closer genetic proximity between the freshwater and estuarine populations (Table 4d) than between the freshwater and salt-marsh populations (Table 4c) (see Results). In addition, the reservoir Lac St Jean (site 15) contained a unique haplotype that was more closely related to those in the upstream estuarine populations (sites 5–7) than to the haplotype in the nearby Saguenay River (site 16), which was closely related to the salt-marsh haplotypes (Fig. 3). However, we cannot completely exclude the possibility that the salt-marsh populations gave rise to the freshwater populations, because several haplotypes were shared among all populations in the Atlantic clade.

The results above are intriguing, as they indicate that invasions likely arose from only a subset of populations from the Atlantic clade. Thus, heterogeneity in the native range characterizes not only divergences between the Atlantic and North Atlantic clades, but also differentiation within the Atlantic clade. Why would invasions arise from particular habitats? Results from another study revealed physiological differences between populations from the Atlantic and North Atlantic clades, with important implications for the ability to invade fresh water (Skelly *et al*. submitted). Such functional and physiological differences might also exist among populations within the Atlantic clade. Differences in selection regime among microhabitats might have led to physiological differences among populations.

Alternatively, patterns of invasions might simply reflect differences in transport opportunity or chance events. The mere location of the Atlantic clade in the upstream estuary might have facilitated its invasion into freshwater habitats. On the other hand, there might be similar probability for either clade to invade, where the clade that invades first would adapt to freshwater conditions, allowing it to competitively exclude the other clade. However, variation in habitat type and physiological differences among populations in the native range do raise questions regarding the role of the selection regime on native populations.

In several other case studies, invasive populations have appeared to emerge from native habitats characterized by disturbance. This pattern appears to hold for invasive ants from the flood plains of Argentina (Tsutsui *et al*. 2001; Caldera *et al*. submitted) and invaders into the Great Lakes from the Ponto-Caspian basin (Lee & Bell 1999; Ricciardi & MacIsaac 2000; May *et al*. 2006). Likewise, the marginal habitats of the invasive clade of *E. affinis* are also characterized by disturbance, with fluctuations in environmental conditions occurring on multiple timescales (seasonal, diel). Fluctuating conditions in the more ancient salt-marsh populations (Table 7) might have selected for quantitative genetic variation for physiological tolerance, enabling the colonization of lower salinity regions of the estuary (Barton & Turelli 1989; Ellner & Hairston 1994; Turelli & Barton 1994). The upstream estuarine populations might have then allowed for pre-adaptation to lower salinities, and served as a stepping-stone for freshwater invasions (Skelly *et al*. submitted).

Environmental disturbance could be reflected in genetic signatures of population bottlenecks and low effective population size (Gelembiuk *et al*. 2006; May *et al*. 2006). For example, invasive populations of zebra mussels most likely arose from the Black Sea basin where N_e, based on Watterson's θ, was estimated at only 80 000 individuals (Gelembiuk *et al*. 2006). However, effective population size of *E. affinis* in the native range was relatively large (female N_e of 10^6–10^7; Tables 3 and 7). Overlapping generations provided by the diapause egg bank might contribute to the large effective population size. As would be expected, effective population size of *E. affinis* was much smaller than the census size, where conservative estimates were on the order of 10^{10} in the middle estuary (Winkler *et al*. 2005).

Evolutionary history in the native range

High genetic divergence between the Atlantic and North Atlantic clades supported that the St Lawrence Middle Estuary constitutes a secondary contact zone between the clades. The timing of population expansion for both clades, estimated from pairwise mismatch analysis of haplotypes (Table 7), greatly preceded the timing of the last glaciation (Wisconsinian glaciation, maximum advance 18 000 years ago) when the St Lawrence drainage basin was covered with ice (Bernatchez 1997). Thus, populations must have persisted in glacial refugia and subsequently colonized the St Lawrence drainage without suffering severe population bottlenecks (Tables 3 and 7). Timing of contact between the two clades in the St Lawrence drainage might have occurred *c*. 7800 years ago after the retreat of the Wisconsinian glaciations, when sea levels and oceanographic conditions in the estuary and Gulf of St Lawrence became comparable to present-day conditions (Fulton & Andrews 1987; Schmidt 1986 in Bernatchez 1997).

Timing of population expansion suggested that population expansion of the Atlantic clade occurred earlier than that of the North Atlantic clade (Table 7). However, the timing is uncertain given the large confidence intervals (Table 7). Also, we cannot rule out the possibility that the timing of the last population expansion event actually represents the timing of the last selective sweep (Bazin *et al*. 2006).

Given the large genetic divergence between Atlantic and North Atlantic clades, it is likely that they occupied separate glacial refugia during the last glacial maximum. The clades probably diverged around the Miocene–Pliocene boundary (~5 million years ago; Lee 2000). Two main regions might have provided glacial refugia for the two clades, one to the southeast and another to the northeast of the St Lawrence basin, as has been proposed for northern fishes (Bernatchez 1997; Lafontaine & Dodson 1997). The current geographic distribution of the Atlantic clade along the east coast of North America (Fig. 1a; Lee 1999, 2000) suggests that the Atlantic clade might have recolonized the St Lawrence basin from the unglaciated coastal waters south and east of the Appalachian Mountains along the Atlantic coastal plains (Dyke & Prest 1987). Colonization could have begun *c*. 12 000 years ago from the south via the Hudson River valley when the brackish Champlain Sea developed (covering the St Lawrence River, Ottawa River and Champlain Lake) (Rubec 1975). In contrast, the North Atlantic clade might have followed an alternative colonization pathway from the northeastern refuge at the Grand Banks near Newfoundland, via the Goldthwait Sea (region of the present St Lawrence Gulf) (Stemberger 1995). This region was characterized by high salinities and cold temperatures (de Vernal *et al*. 1993), but transformed into estuarine conditions *c*. 8000–7500 years ago. The current restricted distribution of the North Atlantic clade in the northern coastal region of the Atlantic (Fig. 1a) might have resulted from colonization arising from this northeastern glacial refugium, following the establishment of the Atlantic clade in the St Lawrence basin and along the Atlantic coast.

Information from the native range: management and policy implications

Our understanding of factors that contribute to the evolution of invasive populations remains poor (Lee 2002). This study revealed a striking degree of heterogeneity in habitat type and genetic structure within the native range of an invasive species. Characterizing the population genetic structure and microhabitat distribution within the native environments is a critical first step toward analysing the invasive potential of populations. Additional information regarding this heterogeneity would provide insights into environmental conditions that tend to give rise to invasive populations. Such information would include: (i) environmental conditions within native habitat types, including information on the selection regimes, and (ii) physiological responses of populations from the native habitats, including quantitative genetic variance underlying the traits critical for survival in the invaded range (e.g. Lee *et al.* 2003, 2007).

Information on the selection regimes in the native range is important because previous selection regimes shape physiological properties, levels of phenotypic plasticity, and the evolutionary potential of invasive populations. For example, anecdotal and qualitative observations suggest that levels of 'disturbance' in the native range might be an important selective agent for generating weedy species (Di Castri 1989). While this property has not been extensively quantified, it is possible that 'disturbance' in the form of temporally fluctuating selection might be an important force for generating invasive populations in some cases. Temporally fluctuating conditions could select for the maintenance of genetic variance for critical phenotypic traits, and promote the evolutionary potential of those traits (Barton & Turelli 1989; Ellner & Hairston 1994; Turelli & Barton 1994). Integrating information on environmental conditions in the native habitats, such as the magnitude and duration of environmental fluctuations, along with physiological responses of the native populations would allow us to discern the types of environmental conditions that might give rise to invasive populations. Such insights would allow us to focus management and mitigation efforts toward those populations that are likely to successfully adapt to novel environments during invasions.

Acknowledgements

This paper was submitted for the international summit on *Microevolutionary Change in Human-altered Environments*, held at the University of California, Los Angeles, 8–11 February 2007. The authors thank Louis Bernatchez and Tom Smith for organizing this symposium. Funding for this research was provided by an NSERC grant to J.J. Dodson and National Science Foundation DEB-0448827 to C.E. Lee. We thank crewmembers of the *RV Mordax* for their assistance in the field. Special thanks are extended to I. St Cyr, B. Laurent, M. Ringuette, D. Kelly and D. Skelly for sample collections, F. Colombani for help with DNA extraction, amplification and sequencing, and F. Lecomte and S. Temblay for helpful discussion. Greg Gelembiuk and Dan Skelly provided insightful editorial comments, and Greg Gelembiuk provided useful methodological suggestions.

References

Anderson DV, Clayton D (1959) *Plankton in Lake Ontario*. Physics Res. Note, no. 1, Div. Res. Dept. Lands and Forests, Maple, Ontario.

d'Anglejan B, Ingram RG, Savard JP (1981) Suspended-sediment exchanges between the St Lawrence estuary and a coastal embayement. *Marine Geology*, **40**, 85–100.

Barton NH, Turelli M (1989) Evolutionary quantitative genetics — how little do we know. *Annual Review of Genetics*, **23**, 337–370.

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.

Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt (*Osmerus mordax*) and their reproductive isolation in the St Lawrence R. estuary (Québec, Canada). *Molecular Ecology*, **6**, 73–83.

Brown J, Idris A (2005) Genetic differentiation of whitefly *Bemisia tabaci* mitochondrial cytochrome oxidase I, and phylogeographic concordance with the coat protein of the plant virus genus *Begomovirus*. *Annals of the Entomological Society of America*, **98**, 827–837.

Burton RS, Lee B (1994) Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. *Proceedings of the National Academy of Sciences, USA*, **91**, 5197–5201.

Caldera EJ, Ross KJ, DeHeer CJ, Shoemaker DD (submitted) Putative native source of the invasive fire ant *Solenopsis invicta* in the USA. *Biological Invasions*.

Caudill CC, Bucklin A (2004) Molecular phylogeography and evolutionary history of the estuarine copepod, *Acartia tonsa*, on the Northwest Atlantic coast. *Hydrobiologia*, **511**, 91–102.

Chu D, Zhang Y-J, Brown JK *et al*. (2006) The introduction of the exotic Q biotype of *Bemisia tabaci* from the Mediterranean region into China on ornamental crops. *Florida Entomologist*, **89**, 168–174.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1675–1660.

Colwell RR (2004) Infectious disease and environment: cholera as a paradigm for waterborne disease. *International Microbiology*, **7**, 285–289.

Di Castri F (1989) History of biological invasions with special emphasis on the Old World. In: *Biological Invasions: A Global Perspective* (eds Drake JA, Mooney HA, di Castri F *et al*.), pp. 1–30. John Wiley, Chichester, UK.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie Physique et Quaternaire*, **41**, 237–263.

Ellner S, Hairston NG (1994) Role of overlapping generations in maintaining genetic-variation in a fluctuating environment. *American Naturalist*, **143**, 403–417.

Engel RA (1952) *Eurytemora affinis*, a calanoid copepod new to Lake Erie. *Ohio Journal of Science*, **62**, 252.

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

Faber DJ, Jermolajev EG, Kossiakina EG (1966) A new copepod genus in the plankton of the Great Lakes. *Limnology and Oceanography*, **11**, 301–303.

Folmer O, Black M, Hoeh W, Lutz R, Vriyenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fulton RJ, Andrews JT (1987) The Laurentide ice sheet. *Géographie Physique et Quaternaire*, **41**, 328p.

Gelembiuk GW, May GE, Lee CE (2006) Phylogeography and systematics of zebra mussels and related species. *Molecular Ecology*, **15**, 1033–1050.

Gillespie J (2001) Is the population size of a species relevant to its evolution? *Evolution*, **55**, 2161–2169.

Goetze E (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 2321–2331.

Heidelberg JF, Heidelberg KB, Colwell RR (2002) Bacteria of the γ-subclass *Proteobacteria* associated with zooplankton in Chesapeake Bay. *Applied and Environmental Microbiology*, **68**, 5498–5507.

Hoelzel AR, Green A (1992) Analysis of population-level variation by sequencing PCR-amplified DNA. In: *Molecular Genetic Analysis of Populations: a Practical Approach* (ed. Hoelzel AR), pp. 159–187. IRL Press, Oxford, UK.

Huq A, Small EB, West PA *et al*. (1983) Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Applied Environmental Microbiology*, **45**, 275–283.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 2257–2263.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Lafontaine P, Dodson JJ (1997) Intraspecific genetic structure of white sucker (*Catostomus commersoni*) in northeastern North America as revealed by mitochondrial DNA polymorphism. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 555–565.

Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, **53**, 1423–1434.

Lee CE (2000) Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate 'populations.'. *Evolution*, **54**, 2014–2027.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology &Evolution*, **17**, 386–391.

Lee CE, Bell MA (1999) Causes and consequences of recent freshwater invasions by saltwater animals. *Trends in Ecology & Evolution*, **14**, 284–288.

Lee CE, Frost BW (2002) Morphological Stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia*, **480**, 111–128.

Lee CE, Remfert JL, Gelembiuk GW (2003) Evolution of physiological tolerance and performance during freshwater invasions. *Integrative and Comparative Biology*, **43**, 439–449.

Lee CE, Remfert JL, Chang Y-M (2007) Response to selection and evolvability of invasive populations. *Genetica*, **129**, 179–192.

May GE, Gelembiuk GW, Panov VE, Orlova MI, Lee CE (2006) Molecular ecology of zebra mussel invasions. *Molecular Ecology*, **15**, 1021–1031.

Meusnier I, Valero M, Olsen J, Stam W (2004) Analysis of rDNA ITS1 indels in *Caulerpa taxifolia* (Chlorophyta) supports a derived, incipient species status for the invasive strain. *European Journal of Phycology*, **39**, 83–92.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Patalas K (1972) Crustacean plankton and the eutrophication of St Lawrence Great Lakes. *Journal of the Fisheries Research Board of Canada*, **29**, 1451–1462.

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

Ricciardi A, MacIsaac HJ (2000) Recent mass invasion of the North American Great Lakes by Ponto-Caspian species. *Trends in Ecology and Evolution*, **15**, 62–65.

Ross KG, Krieger MJB, Keller L, Shoemaker DD (in press) Genetic variation and structure in native populations of the fire ant *Solenopsis invicta*: Evolutionary and demographic implications. *Biological Journal of the Linnean Society*.

Roy S (2002) Effet des marées sur la dynamique des sédiments de surface dans la zone de turbidité maximum du fleuve Saint-Laurent. Honour project, Département de Biologie, Université Laval, Quebec.

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

Rubec PJ (1975) Fish distribution in Gatineau Park, Québec, in relation to postglacial dispersal, man's influence, and eutrophication. *Canadian Field-Naturalist*, **89**, 389–399.

Saltonstall K (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences, USA*, **99**, 2445–2449.

Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Simons RD, Monismith SG, Johnson LE, Winkler G, Saucier FJ (2006) Zooplankton retention in the estuarine transition zone of the St Lawrence Estuary. *Limnology and Oceanography*, **51**, 2621–2631.

Skelly D, Chau FC, Winkler G, Chang Y-M, Lee CE (Submitted) Limits to range expansions into freshwater habit: Physiological contrasts between sympatric invasive and non-invasive populations of the copepod. *Eurytemora Affinis*.

Stemberger RS (1995) Pleistocene refuge areas and postglacial dispersal of copepods of the northeastern United States. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 2197–2210.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2001) Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Molecular Ecology*, **10**, 2151–2161.

Turelli M, Barton NH (1994) Genetic and statistical analyses of strong selection on polygenic traits — what, me normal. *Genetics*, **138**, 913–941.

de Vernal A, Guiot J, Turon J-L (1993) Late and post-glacial paleoenvironments of the Gulf of St Lawrence: marine and terrestrial palynological evidence. *Géographie Physique et Quaternaire*, **47**, 167–180.

Vincent WF, Dodson JJ, Bertrand N, Frenette JJ (1996) Photosynthetic and bacterial production gradients in a larval fish nursery: the St Lawrence River transition zone. *Marine Ecology Progress Series*, **139**, 227–238.

Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Willey A (1923) Distribution of Copepoda in Canadian waters. *Contributions to Canadian Biology* (new series), **1**, 303–334.

Winkler G, Dodson JJ, Bertrand N, Thivierge D, Vincent WF (2003) Trophic coupling across the St Lawrence River estuarine transition zone. *Marine Ecology Progress Series*, **251**, 59–73.

Winkler G, Sirois P, Johnson LE, Dodson JJ (2005) Invasion of an estuarine transition zone by *Dreissena polymorpha* veligers had no detectable effect on zooplankton community structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 578–592.

Gesche Winkler studies the ecology of zooplankton communities in estuarine transition zones. She is interested in processes structuring biological diversity, including evolutionary, trophodynamic and environmental aspects. Julian J. Dodson's key research themes encompass the evolutionary ecology and phylogeography of fishes and crustaceans, as well as the adaptive significance and evolution of alternative life-history strategies in fishes. Carol E. Lee's research focuses on evolutionary mechanisms that allow populations to breach boundaries between habitat (niche evolution). Her research includes the population genetics, evolutionary physiology, and genomics of invasive species.