Molecular Ecology (2006) **15**, 1033–1050 doi: 10.1111/j.1365-294X.2006.02816.x

Phylogeography and systematics of zebra mussels and related species

GREGORY W. GELEMBIUK, GEMMA E. MAY and CAROL EUNMI LEE
Wisconsin Institute of Rapid Evolution, Department of Zoology, 430 Lincoln Drive, University of Wisconsin-Madison, Madison, WI 53706, USA

Abstract

The genus *Dreissena* includes two widespread and aggressive aquatic invaders, the zebra mussel, *Dreissena polymorpha*, and the quagga mussel, *Dreissena bugensis*. This genus evolved in the Ponto-Caspian Sea basin, characterized by dynamic instability over multiple timescales and a unique evolutionary environment that may predispose to invasiveness. The objectives of this study were to gain insights into the demographic history of *Dreissena* species in their endemic range, to reconstruct intraspecific phylogeographic relationships among populations, and to clarify systematics of the genus, using DNA sequences from the mitochondrial cytochrome oxidase I (COI) gene. We found four deeply diverged clades within this genus, with a basal split that approximately coincided with the Cretaceous–Tertiary boundary. Divergence events within the four base clades were much more recent, corresponding to geographically disjunct sets of populations, which might represent species complexes. Across all taxa, populations of *Dreissena* shared a common pattern of genetic signatures indicating historical population bottlenecks and expansions. Haplotype diversity was relatively low in Ponto-Caspian drainages relative to more stable tectonic lakes in Greece, Macedonia, and Turkey. The phylogeographic and demographic patterns in the endemic range of *Dreissena* might have resulted from vicariance events, habitat instability, and the high fecundity and passive dispersal of these organisms.

Keywords: Anatolia, bivalve, demography, *Dreissena*, fluctuating environment, tectonic lakes, Turkey

Received 30 March 2005; revision received 10 October 2005; accepted 31 October 2005

Introduction

The genus *Dreissena*, which includes the invasive zebra mussel, *Dreissena polymorpha*, and quagga mussel, *Dreissena bugensis*, evolved in the Ponto-Caspian Seas and the preceding Paratethys Sea. Strikingly, a very high proportion (70%) of species that have invaded the Laurentian Great Lakes between 1985 and 2000 are endemic to the Ponto-Caspian Seas (Ricciardi & MacIsaac 2000). The Ponto-Caspian region is characterized by dynamic instability on multiple timescales and constitutes a unique evolutionary environment, perhaps predisposing to invasiveness. This region has peculiar characteristics in terms of species composition. Although the brackish waters of the Ponto-Caspian are depauperate in species, they do contain many endemics arising from the ancient Paratethys Sea. This endemic fauna is drawn from a limited number of families and includes a set of species flocks. In particular, *Dreissena* constitutes an abundant and widespread species flock in this region, composed of species and subspecies occupying diverse ecological niches (Kinzelbach 1992; Starobogatov 1994). Thus, the objectives of this study were to (i) examine effects of geographic and demographic history on genetic signatures of *Dreissena* populations in their endemic range and (ii) to clarify systematic relationships within the genus *Dreissena* on a global scale.

The genetic structure of *Dreissena* populations has never been extensively examined in the endemic Ponto-Caspian range. We anticipated that specific geological and climatic events, including fluctuations in water level and interconnectedness between the Ponto-Caspian Seas, would have shaped intraspecific genetic patterns. In addition, passive dispersal in this genus would have led to a high degree of genetic fragmentation among drainages. The fluctuations and instability of the Ponto-Caspian Sea basin, in combination

Correspondence: Carol Eunmi Lee, Fax: (608) 265 6320; E-mail: carollee@wisc.edu

with passive dispersal and an inability to escape environmental shifts, would have led to frequent extinctions and colonizations and favoured high reproductive potential (Stearns 1992; Henle *et al*. 2004). Frequent extinction/recolonization events, coupled with a high intrinsic population growth rate after colonization, would have generated genetic signals of population bottlenecks and expansions. In contrast to the fluctuations of the Ponto-Caspian Seas, populations in ancient stable lakes, such as Lakes Prespa and Ohrid in the Balkans (Stankoviç 1960; Watzin *et al*. 2002), should have experienced relatively static environments. Such conditions would have resulted in genetic signatures reflecting greater stability, such as higher haplotype diversity. These genetic signatures might yield information on the evolutionary regimes imposed by the Ponto-Caspian Sea region, a major source for generating invasive populations.

Our second objective was to clarify the systematics and taxonomy of *Dreissena*, because this genus contains many described species and subspecies for which systematic and phylogeographic relationships are unclear (Locard 1893; Kinzelbach 1992; Schütt 1993; Rosenberg & Ludyanskiy 1994). Uncertainties in systematic designations could result in unsound hypotheses or incorrect conclusions on evolutionary and invasion history of populations. For example, ancient freshwater populations in Eurasian lakes have been proposed as potential sources of invasions (Strayer 1999) but their phylogenetic relationship to invasive populations of *D. polymorpha* has been uncertain. Reconstructions based on geography, morphology, and the fossil record have been compromised by morphological variation and overlap in shell morphology among species (Andrusov 1897; Babak 1983; Nuttall 1990; Kinzelbach 1992; Schütt 1993; Starobogatov 1994; Claxton *et al*. 1998), possibly as a result of environmental plasticity. Consequently, the number of described species ranges from about 4 to over 20 with various subspecies designations (Andrusov 1897; Nuttall 1990; Kinzelbach 1992; Schütt 1993; Rosenberg & Ludyanskiy 1994; Starobogatov 1994).

In order to examine the evolutionary history of dreissenid species, we reconstructed intraspecific phylogeographic relationships and analysed patterns of haplotype frequencies across a broad geographic range in the endemic region, including the Ponto-Caspian basin and lakes in Greece, Macedonia, and Turkey. Coalescent analysis was used to make inferences about population size, fragmentation, colonization, and migration. To clarify systematic relationships, we constructed a comprehensive phylogeny of *Dreissena* including described species of the genus *Dreissena* and *D. polymorpha* populations of uncertain systematic status from Turkey. Our analyses employed DNA sequences (606 bp) from the mitochondrial COI gene. Previous molecular systematic and phylogeographic studies of this genus had been more limited in geographic scope (Therriault *et al*. 2004). Comparing invasive and non-invasive populations of *Dreissena* might provide insights into associations between source habitats and the propensity to invade. It is conceivable that adaptation to unstable environments contributed to the invasive potential of Ponto-Caspian species (Geary *et al*. 2000; Ricciardi & MacIsaac 2000; Reid & Orlova 2002).

Materials and methods

Population sampling, DNA extraction, and COI DNA sequencing

Sampling design, including sample sizes and locations, and methods of DNA extraction and sequencing are discussed in the accompanying paper (see Methods, Table 1 and Fig. 1 in May *et al*. submitted). See Fig. 4 for photographs of taxa and Appendix S2 in Supplementary materials for physical descriptions.

Haplotype networks

Haplotype networks were constructed using the statistical parsimony software package TCS 1.13 (Clement *et al*. 2000). These networks were used to represent relatedness among extant haplotypes in *Dreissena polymorpha* (sites 9, 11–12, 15–18, 22–28), *Dreissena caputlacus* (site 27), and *Dreissena stankovici* (sites 20, 21).

Tests of population expansion

Analyses of population growth were performed for individual populations of *D. stankovici* and *D. caputlacus*, and for the collective population set in *D. p. polymorpha* and *D. p. anatolica*. For the latter two species, analysis was not performed on individual populations because of prohibitively low genetic polymorphism. DNASP 4.0 (Rozas & Rozas 1999) was used to estimate Fu and Li's *D* statistic (Fu & Li 1993) to test for deviations from a neutral Wright–Fisher model consistent with population expansion. The coalescent analysis program FLUCTUATE 1.5 (Kuhner *et al*. 1998) was used to examine the joint likelihood surface for θ ($\theta = N_e\mu$ for mtDNA, where μ = mutation rate and N_e = effective population size, provided $N_e = 2N_f$) and population growth rate g under a model of exponential growth. Joint estimation is required in this case because the parameters are not independent — a higher growth rate g would result in a larger value of θ.

The program ARLEQUIN 2.001 (Schneider *et al*. 2000) was used to estimate τ ($2\mu t$, where t = time in generations and μ = mutation rate/generation), the time since demographic expansion, as well as initial and final θ, under a model of sudden demographic expansion (Schneider & Excoffier 1999). A generalized least squares approach is employed for parameter fitting to the pairwise mismatch distribution.

This analysis was performed for individual populations of *D. stankovici* and *D. caputlacus*. Alternative methods were used when a model of range expansion appeared more suitable than one of simple demographic expansion, or when mandated by technical difficulties in implementation (see below). Validity of the sudden expansion model was assessed using parametric bootstrap coalescent simulations [goodness-of-fit test based on proportion of bootstrap replicates in which the sum-of-squared-deviations (SSD) for the simulated data sets equalled or exceeded the observed SSD]. ARLEQUIN parameter estimation algorithms failed for one population (Lake Prespa). Consequently, these data were analysed following the methods of Schneider & Excoffier (1999) as in ARLEQUIN, but using alternative nonlinear optimization algorithms (Kelley 1998), with coalescent simulations performed using TREEVOLVE (Grassly & Rambaut 2003).

A four-parameter model of sudden range expansion (Excoffier 2004) was fitted for *D. caputlacus* and *D. p. polymorpha* mismatch distributions with a generalized least squares approach using nonlinear optimization algorithms (Kelley 1998). Under this model, a single deme (size θ_0) instantaneously colonizes, at time τ ($2\mu t$), an infinite number of demes (each of size θ_1), which exchange migrants at rate M ($M = 2Nm = \theta_1 m/\mu$). Under such spatial expansion and a relatively 'low' migration rate (i.e. $Nm < 50$) between colonized demes, a large number of coalescent events occur within the individual colonized demes. This results in a pairwise mismatch peak at 0 mismatches (sequence identity) before gene lineages enter the overall population pool (note that coalescent analysis looks backward in time). The latter coalescent events generate a second peak corresponding to time of onset of expansion. For *D. p. polymorpha* this spatial expansion model was fitted to the mean intrapopulation mismatch distribution (averaged across all sampled populations).

Technical limitations prevented appropriate application of least squares fitting of expansion models to the *D. p. anatolica* pairwise mismatch distribution. As an alternative methodology, parameters of a model of sudden demographic expansion (τ, initial θ, and final θ) were estimated using the program SITES (Wakeley & Hey 1997). SITES employs information from the number of sites segregating at particular frequencies in the sample, fitting observed values to the expectation under particular expansion parameters. This analysis was performed by pooling haplotypes from all populations sampled, since gross predominance of haplotype F (see below) rendered individual sites largely uninformative. Tests of population expansion have generally been viewed as robust to population substructure, since pooling of demes under equilibrium island models would not result in an artefactual signal of expansion. However, historical branching processes resulting in fine-scale population structure can produce genetic signatures similar to population expansion when pooled samples are used, due to accumulation of new mutations in isolated demes (Ptak & Przeworski 2002). We cannot exclude such a possibility here.

Intraspecific analysis of population structure

The Bayesian Markov chain Monte Carlo (MCMC) program BATWING 1.02 (Wilson *et al*. 2003) was used to conduct coalescent analysis under a bifurcating population model to estimate (i) posterior probabilities for alternative rooted phylogenetic trees of populations, and (ii) posterior probabilities of haplotype ancestry, rooting intraspecific haplotype networks. Specifically, a rooted population tree was estimated for the *D. polymorpha* species complex, and ancestral haplotypes were inferred for *D. polymorpha*, *D. stankovici*, and *D. caputlacus*. Total population size was assumed constant over time (a constraint required for satisfactory mixing of the MCMC chain), with sequence evolution occurring under a unique event polymorphism model. Truncated uniform priors were used for θ and times of population divergence, while a uniform Dirichlet (1,1, … ,1) prior was used for relative subpopulation size. MCMC samples from the posterior probability distribution of rooted topologies were analysed using the MRBAYES 3.01 sumt function (Huelsenbeck *et al*. 2001), generating a consensus tree.

Branch lengths of the population tree (specified in terms of $\boldsymbol{T} = \mu t$, or expected number of mutations along each branch) and values of θ, reflecting population size, were estimated using the Bayesian program MCMCCOAL (Rannala & Yang 2003). These parameters were estimated under a Hasegawa–Kishino–Yano (HKY) model of sequence evolution, conditional on the topology obtained using BATWING. MCMCCOAL requires priors for θ for all current and ancestral populations, and priors for branch lengths in the population tree. Priors were generated via an empirical Bayes approach. An identical prior for θ was used for each population. The two parameters of this gamma prior were obtained by calculating the mean and variance of Watterson's θ (1975) across all sampled populations. A gamma prior for branch length ($\boldsymbol{T}$) was likewise obtained from the mean and variance for branch lengths in a genetic distance-based Kitsch population tree. The Kitsch tree, which should approximate the coalescent analysis-based tree, was constructed using PHYLIP (Felsenstein 1998), employing corrected average pairwise Tamura and Nei distances between populations (calculated using ARLEQUIN). Posterior means of MCMC samples from MCMCCOAL were used as parameter estimates.

The Bayesian MCMC program IM (Hey & Nielsen 2005) was used to simultaneously estimate divergence time and migration rate between pairs of populations, as well as values of θ for the ancestral and each daughter population,

under an HKY model of sequence evolution (the most appropriate mutational model available in IM). Methods that assume strict population bifurcation without migration, or the inverse, conflate divergence time with effects of subsequent migration. Truncated uniform priors were used for all parameters. In accordance with author's recommendation (Hey & Nielsen 2005), the modal value of the integrated likelihood surface was used as the parameter estimate.

Coalescent population parameters such as θ and T entangle demographic parameters of interest with the mutation rate μ. Demographic parameters were extracted by assuming a mutation rate of 0.5%/million years (Myr), based on rates of COI sequence divergence in arcid bivalves with calibration from the fossil record (Marko 2002). Specifically, this mutation rate (0.5%/Myr) represents a mean across three fossil calibration points, with lower and upper bounds as follows (0.34, 0.48; 0.54, 0.61; 0.38, 0.53) in percentage per million years (Marko 2002). Marko's (2002) calibration was selected based on its relative rigour (first appearance in the fossil record of three geminate species pairs) and its similarity to estimates of mitochondrial divergence rates in numerous other marine species. The uncertainty inherently associated with the use of such a calibration would represent an additional source of error in our time estimates (see next section).

Phylogenetic reconstruction

A phylogeny of the genus *Dreissena* was constructed using the COI gene. Sequences were aligned using CLUSTAL V (Higgins *et al*. 1992). Bayesian inference of the phylogeny was performed using MRBAYES 3.01 (Huelsenbeck *et al*. 2001) via MCMC sampling for 5 million generations (four simultaneous MC chains; sample frequency 100; burn-in 500 000 generations; chain temperature 0.2) under a codon-structured GTR + G model with 4 gamma rate categories (a parameter-rich model in keeping with the recommendations of Huelsenbeck & Rannala 2004 for Bayesian phylogenetic inference). A favoured tree was obtained by majority rule consensus.

Additional phylogenetic reconstructions using maximum-parsimony, distance matrix (neighbour joining), and maximum-likelihood (ML) approaches were performed using PAUP* 4.0 (Swofford 1998). For neighbour-joining and ML reconstructions, MODELTEST (Posada & Crandall 1998) was used to select the substitution model (using bottom-up heirarchical likelihood-ratio tests) and estimate base frequencies, rate matrix, and gamma shape parameter that best fit the data. The reconstructions were performed using heuristic searches with tree-bisection–reconnection (TBR) branch swapping. To estimate statistical support for tree topology, 100 bootstrap replicates were performed. For all phylogenetic reconstructions, *Mytilopsis leucophaeata* and *Congeria kusceri* sequences were used as outgroup species (GenBank Accession nos UU47649, AF325444).

A parsimony tree was constructed with 10 replicates performed within each heuristic search, using random taxon addition. A neighbour-joining tree was constructed using distances estimated by ML, under a K81uf + G substitution model (Kimura 3-parameter with unequal base frequencies incorporating rate heterogeneity) and gamma shape parameter 0.1930. The ML tree reconstructions used substitution model K81uf + G and gamma shape parameter 0.1802, and employed a reduced data set because of lengthy computation (haplotypes enumerated in legend of Fig. 2).

Molecular clock analyses were performed to date the primary divergence points in *Dreissena*. An ML tree was estimated using PAUP* 4.0 using the K81uf + G model and the reduced data set (enumerated in Fig. 2 legend) without a molecular clock constraint, and then applying such a constraint. A likelihood-ratio test comparing the unconstrained tree with the constrained tree with identical topology did not reject the hypothesis of a global molecular clock ($P > 0.1$). MRBAYES 3.01 was also used to perform molecular clock analysis under a birth-death prior, using the codon-structured GTR + G model. Only MCMC tree samples that respected topological constraints imposed by the four primary clades were used for this analysis. Additional analyses were performed to further corroborate the approximate validity of dates estimated using the clock of Marko (2002) and COI sequence data, and to provide a rough lower bound on the date of the basal *Dreissena* divergence event (see Appendix S1 in Supplementary materials).

Results

Intraspecific genetic polymorphism

Haplotype relatedness for the *Dreissena polymorpha* species complex, *Dreissena caputlacus*, and *Dreissena stankovici* was represented using statistical parsimony networks (Fig. 1). For the *D. polymorpha* complex, two clusters were apparent, centred on haplotypes B and F. Haplotype F, central in the *D. polymorpha anatolica* cluster, was five mutational steps removed from both B (central to the *D. polymorpha polymorpha* cluster) and E (*D. polymorpha gallandi*), while B and E were separated by two steps (Fig. 1A). Under coalescent expectations, the connectedness and high frequency of B suggested that it was basal for *D. p. polymorpha* (Crandall & Templeton 1993). Coalescent simulations under a unique event polymorphism model of mutation yielded a 76% posterior probability that B was ancestral for *D. p. polymorpha*. The inferred haplotype labelled Ω_1 (Fig. 1A) situated between B and E, exhibited the highest posterior probability (43%) of being ancestral for the entire *D. polymorpha* species complex. For the *D. caputlacus*

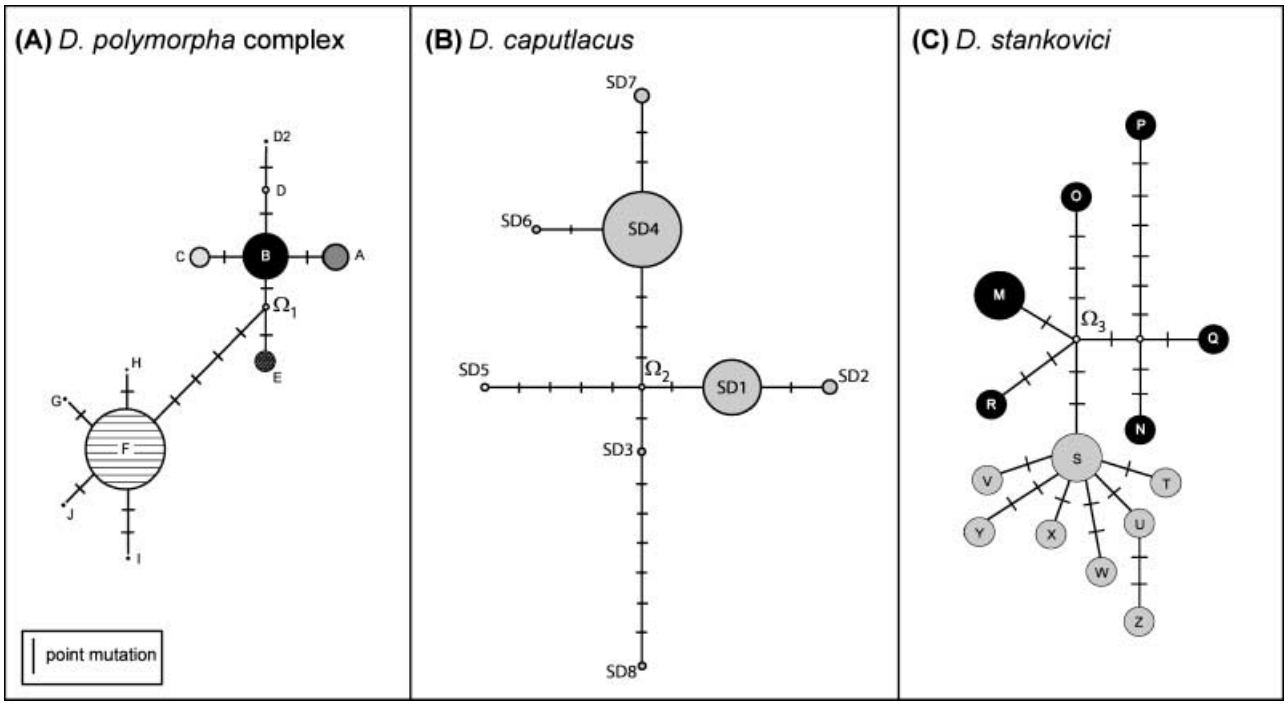

Fig. 1 Haplotype networks of species of *Dreissena*. Point mutations are indicated by dashes (—). (A) *Dreissena. polymorpha* complex. Circle area reflects haplotype frequencies in samples collected from the endemic range of *D. polymorpha* (sites 9, 11, 15–18, 22–28). Haplotype patterns correspond to those in Fig. 2. Ω_1 represents the most probable ancestral root for this complex. (B) *Dreissena caputlacus*. Circle area reflects haplotype frequencies in the sample collected from Seyhan Dam, Turkey (site 27). Ω_2 represents the most probable ancestral root. (C) *Dreissena stankovici*. Circle area reflects the frequencies of haplotypes in samples collected from Lakes Prespa and Ohrid. The lighter circles represent haplotypes from Lake Prespa (site 21), while the darker circles represent haplotypes from Lake Ohrid (site 20). Ω_3 represents the most probable ancestral root.

haplotype network, the inferred haplotype Ω_2 exhibited the highest root probability (39%) (Fig. 1B). For *D. stankovici*, Lakes Ohrid and Prespa formed distinct haplogroups containing mutually exclusive subsets of haplotypes. Coalescent analysis of possible *D. stankovici* ancestral roots suggested that the inferred haplotype Ω_3, situated central to the Lake Ohrid clade (Fig. 1C), might have been ancestral (59% posterior probability). In phylogenetic reconstruction of the genus, all Lake Ohrid haplotypes were basal within *D. stankovici* (although degree of support varied among methods) (Fig. 2), supporting placement of the *D. stankovici* ancestral root within Lake Ohrid. For Lake Prespa, haplotype S was most likely ancestral (63% posterior probability).

Evidence of population expansions

Populations showed evidence of a history of expansion events. Haplotype networks exhibited starlike phylogenies (Fig. 1), a pattern consistent with population growth (Slatkin & Hudson 1991). Tests for population growth were performed for individual populations of *D. stankovici* and *D. caputlacus*, and for the collective population set in *D. p. polymorpha* and *D. p. anatolica*, using FLUCTUATE (Kuhner *et al*. 1998). Stable estimates and substantial positive values of g were obtained for Lake Ohrid *D. stankovici*, *D. caputlacus*, *D. p. polymorpha*, and *D. p. anatolica*, with the 99% confidence interval excluding a value of 0 for both *D. p. polymorpha* and *D. p. anatolica*. Although the starlike phylogeny of the Lake Prespa population contained little information on θ, precluding stable joint estimation of growth rate and θ, high values were favoured for both parameters (i.e. likelihood maximized at the high edge of the permitted parameter space). Since estimates from FLUCTUATE are known to suffer from a degree of upward bias (i.e. potentially comparable to the magnitude of the growth estimate for *D. caputlacus*) Fu and Li's *D* statistic (Fu & Li 1993) was also calculated to test for deviations from a neutral Wright–Fisher model consistent with population expansion (Table 1). This statistic generally appeared to support inferences of population expansion.

History of expansion events

To further characterize apparent expansions, a model of sudden demographic growth was fit to the pairwise sequence mismatch distribution. A model of simple demographic expansion was applied to *D. stankovici* from

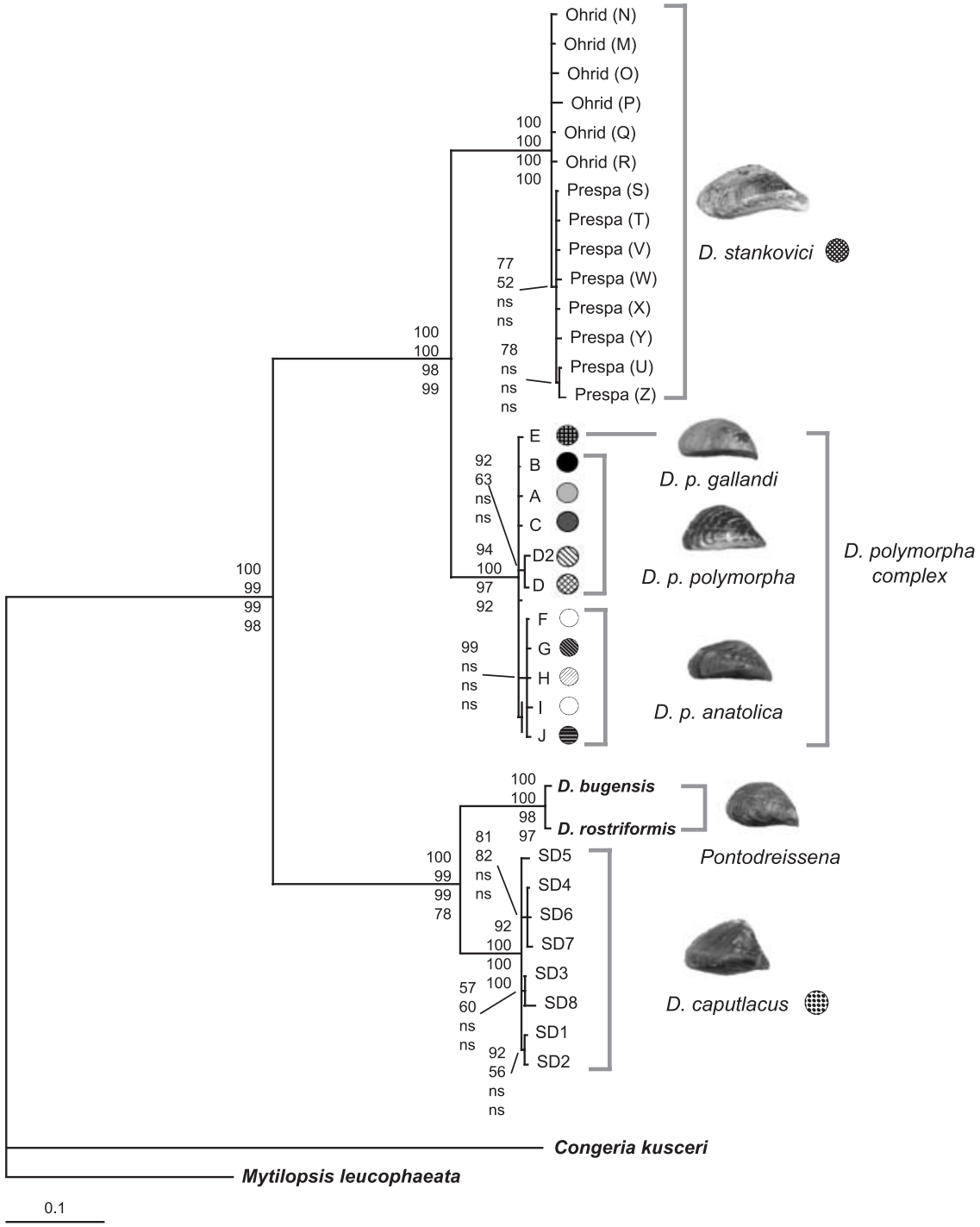

Fig. 2 Phylogeny of the genus *Dreissena* using sequences (606 bp) of the mitochondrial cytochrome oxidase I (COI) gene. For the *Dreissena polymorpha* complex, distinct haplotypes are shown at branch tips (see Fig. 2 in May *et al*. submitted for locations of haplotypes), while for *Dreissena stankovici* and *Dreissena caputlacus* sampling locations are shown. Patterned circles correspond to those in Fig. 2 of May *et al*. (submitted). The phylogeny shown was constructed using a Bayesian approach (see Materials and methods). Scale bar corresponds to number of substitutions per site. Numbers next to nodes are posterior probabilities (top number), and bootstrap values based on 100 bootstrap replicates using parsimony (second number), distance (third number), and maximum-likelihood (fourth number) reconstructions. The maximum-likelihood reconstructions employed a reduced data set because of lengthy computation, and included: *D. polymorpha* complex haplotypes A, D2, E, F, and H; *D. stankovici* haplotypes T, R, Z, Q, and P; *D. caputlacus* haplotypes SD1, SD4, and SD7; *Dreissena bugensis*, and *Dreissena rostriformis*. Nonsignificant bootstrap values are indicated by ns (either due to clade support < 50%, or lack of estimation due to use of a restricted data set in the case of maximum likelihood).

Table 1 Statistical tests of population expansion

Population	Fu and Li's *D* (*P* value)	FLUCTUATE: *g* [$\theta_{per\ site}$]
Lake Prespa *D. stankovici*	−2.614 (0.005)**	496 [0.01][a]
Lake Ohrid *D. stankovici*	−2.311 (0.007)**	663 [0.0871]
Seyhan *D. caputlacus*	−2.198 (0.027)*	155 [0.0116]
D. p. polymorpha	−0.296 (0.383)	3622** [0.0099]
D. p. anatolica	−3.166 (0.003)**	3198** [0.0125]

All *P* values are given as independent hypotheses (non-Bonferroni adjusted).
*Significant at 0.05 level.
**Significant at 0.01 level.
[a]θ fixed at 0.01 to obtain stable estimate of *g*.

Lakes Prespa and Ohrid and *D. caputlacus* from Seyhan Dam (Table 2). In each case, the fit of the model of sudden expansion appeared adequate (i.e. parametric bootstrap goodness-of-fit tests did not reject the model). Estimated times of these expansion events are shown in Table 3. In the case of *D. caputlacus*, although the simple model of sudden demographic expansion was not rejected in a test of goodness-of-fit, the pairwise mismatch distribution exhibited a prominent peak at 0 mismatches (corresponding to full sequence identity) in addition to the primary peak at 5 mismatches. Such a pattern would be expected given range expansion (as opposed to purely demographic expansion) with a relatively 'low' migration rate (i.e. $Nm < 50$) between colonized demes (Excoffier 2004). Compared to a model of purely demographic expansion, a four-parameter model of sudden range expansion resulted in a substantially better model fit for *D. caputlacus* (sum of squared deviations reduced from 0.0540 to 0.0262; parameter estimates, with θ specified per sequence, $\theta_{source\_deme} = 0.0029$, $\theta_{colonized\_demes} = 0.949$; $\tau = 4.44$; $M = 2.68$). This estimate of τ would correspond to an onset of expansion *c.* 740 000 years before present (BP). A model of range expansion was also applied to *D. p. polymorpha*, given that the *D. p. polymorpha* data set was composed of multiple partially differentiated demes (rather than a single panmictic population as assumed under the simple demographic expansion model). The time of onset of expansion for *D. p. polymorpha* was estimated as $\tau = 0.65$, corresponding to an age of 110 000 years, with an estimate of $\theta_{source\_deme} = 0.0028$ (model fit was insensitive to the value of *M* and the magnitude of $\theta_{colonized\_demes}$ was too large to permit accurate estimation). Technical limitations prevented appropriate application of least squares model fitting to the *D. p. anatolica* pairwise mismatch distribution. The program SITES (Wakeley & Hey 1997) was used as an alternative method of estimation of τ for *D. p. anatolica* under a model of simple demographic expansion. This provided a value of $\tau = 0.15$, corresponding to an age of 25 000 years. Although SITES implicitly requires an assumption of panmixia, model fit to the *D. p. anatolica* data set appeared adequate.

Phylogenetic analysis within and among species

Phylogenetic reconstruction of the genus. The *Dreissena* populations we sampled fell into four deeply diverged clades that formed a symmetric phylogenetic tree (Fig. 2). Turkish populations of *D. polymorpha* formed a highly supported clade (94% posterior probability, 92–100% bootstrap values; Fig. 2) with endemic and invasive populations of *D. p. polymorpha*. *D. stankovici* formed a highly supported clade (100% posterior probability, 98–100% bootstrap values) with haplotypes of the *D. polymorpha* complex [*D. p. polymorpha* plus *D. polymorpha* found in Turkey (sites 22–28)]. Populations of *D. stankovici* from Lakes Prespa and Ohrid contained many haplotypes that formed a clade (100% posterior probability) where all Lake Ohrid haplotypes were basal (Fig. 2). *D. caputlacus* contained eight haplotypes, and formed a clade (100% posterior probability, 78–99% bootstrap values) with the pontodreissenids *D. bugensis* and *D. rostriformis*. Populations of *D. bugensis* contained only one haplotype, which formed a clade with *D. rostriformis* (100% posterior probability, 97–100% bootstrap values) with little divergence (haplotypes separated by 3 point mutations) (Fig. 2). DNA sequences from *D. rostriformis distincta* and *D. rostriformis compressa* were identical. Populations and species of *Dreissena* currently known to be highly invasive (*D. p. polymorpha* haplotypes A and B, and *D. bugensis*) were not monophyletic (Fig. 2), raising the possibility that the propensity to invade might have arisen multiple times independently.

Table 2 Mismatch distribution analysis of *Dreissena stankovici* and *Dreissena caputlacus* populations

	Lake Prespa *D. stankovici*	Lake Ohrid *D. stankovici*	Seyhan Dam *D. caputlacus*
τ (95% CI)	1.982 (1.024, 4.504)	3.767 (1.556, 7.527)	5.496 (1.844, 10.637)
θ_0 (95% CI)†	0.232 (0.000, 0.600)	0.989 (0.000, 4.756)	0.005 (0.000, 1.586)
θ_1 (95% CI)†	1126 (2.960, 6965)	45.001 (15.509, 6965)	5.290 (1.677, 1359.04)
$P_{(Sim.\ Ssd \geq Obs.\ Ssd)}$*	0.735	0.120	0.158

*Test of model goodness-of-fit.
†θ per sequence.

Table 3 Estimated dates of divergence or population expansion, based on the COI molecular clock calibration of Marko (2002)

Divergence event	Timing of divergence (95% CI)
Dreissena basal split:	68 Ma (35–125); 71 Ma*
Carinodreissena + *Dreissena sensu stricto* vs. *Pontodreissena* + *D. caputlacus*	
Dreissena major clades:	21 Ma (10–40); 29 Ma*
Carinodreissena (*D. stankovici*) vs. *Dreissena sensu stricto* (*D. polymorpha*)	
Dreissena major clades:	17 Ma (8–30); 28 Ma*
D. caputlacus vs. *Pontodreissena* (*D. bugensis* + *D. rostriformis*)	
D. stankovici Lake Ohrid vs. Lake Prespa	840 000 BP (400 000–1 600 000)
D. polymorpha subspecies:	730 000 BP (290 000–1 370 000)
D. p. anatolica vs. *D. p. polymorpha* + *D. p. gallandi*	
D. polymorpha subspecies:	510 000 BP (140 000–1 030 000)
D. p. gallandi vs. *D. p. polymorpha*	
D. p. polymorpha Black/Caspian Sea split:	166 000 BP (59 000–783 000)
D. p. polymorpha Black Sea vs. Ural†	
D. rostriformis vs. *D. bugensis*	*c.* 500 000 BP‡
Population expansion events	**Timing of divergence (95% CI)**
D. caputlacus	920 000 BP (310 000–1 780 000)
D. stankovici Lake Ohrid	640 000 BP (270 000–1 290 000)
D. stankovici Lake Prespa	340 000 BP (180 000–770 000)
D. p. polymorpha	*c.* 110 000 BP
D. p. anatolica	*c.* 25 000 BP

*The first value was estimated using Bayesian analysis (under a birth-death prior), while the second was estimated with maximum-likelihood analysis. Bayesian analyses using a uniform prior for the molecular clock produced posterior means that were shifted toward the maximum-likelihood estimates (data not shown).
†This estimated date assumes that the Ural River is representative of Caspian Sea drainages and that introgression from the Black Sea has contaminated areas proximal to the Volga River.
‡High uncertainty due to low information content of samples (reciprocally monophyletic species separated by few mutations).

A molecular clock analysis was used to date the three primary divergence points within *Dreissena* (see Table 3, first three lines). Alternative bivalve fossil divergence points and additional genes (mitochondrial 16S and nuclear 18S) were used to corroborate results obtained using the clock of Marko (2002) with COI. The alternative estimates of the timing of basal divergence in *Dreissena* were roughly comparable and consistent with the date of *c.* 68 million years ago (Ma) (Table S1, Supplementary materials). To provide a lower bound on the date of the basal *Dreissena* divergence event, we used Bayesian phylogenetic analysis anchored with a minimum date for the *Congeria* vs. *Mytilopsis* + *Dreissena* split (lower Eocene boundary, 54 Ma), resulting in an estimated minimum age of 26 Ma (95% CI 14–42) for the *Dreissena* basal split (see Appendix S1 in Supplementary materials).

Intraspecific population structure. Intraspecific population structure and population divergence times were examined using coalescent analysis, so as to incorporate the effects of both mutation and lineage sorting. The topology of the *D. polymorpha* population tree, estimated using the program BATWING (Wilson *et al*. 2003), showed three clades corresponding to the three subspecies, with good support for the *D. p. gallandi* and *D. p. anatolica* clades (Fig. 3). Coalescent analysis provided a rooted tree, and suggested a basal split between *D. p. anatolica* and the other two subspecies (*D. p. polymorpha* + *D. p. gallandi*), although support for a specific rooting was quite low. Divergence times were estimated conditional on this topology. The age of the basal split was estimated as *c.* 730 000 years, while the age of the split between *D. p. gallandi* and *D. p. polymorpha* was estimated as *c.* 510 000 years (Table 3). Under an assumption of one generation per year, the current *D. polymorpha* population sizes implied by the estimated values of θ ranged from 60 000 to 300 000, with a median population size of 110 000.

The program IM (Hey & Nielsen 2005) was used to examine divergence time and migration between Lakes Prespa and Ohrid populations of *D. stankovici*. There appeared to be little evidence of migration following the initial divergence event, with the highest value of marginal posterior density essentially at 0 (lower limit of resolution). The estimated value of $\theta_{per\_sequence}$ was 20.9 (95% CI 10.7–244.5) for Lake Prespa, 92.7 (95% CI 39.0–1171) for Lake Ohrid, and 3.95 (95% CI 0.78–23.7) for the ancestral population, and $\boldsymbol{T}$ was estimated as 2.45 (95% CI 1.16–4.64). The esti-

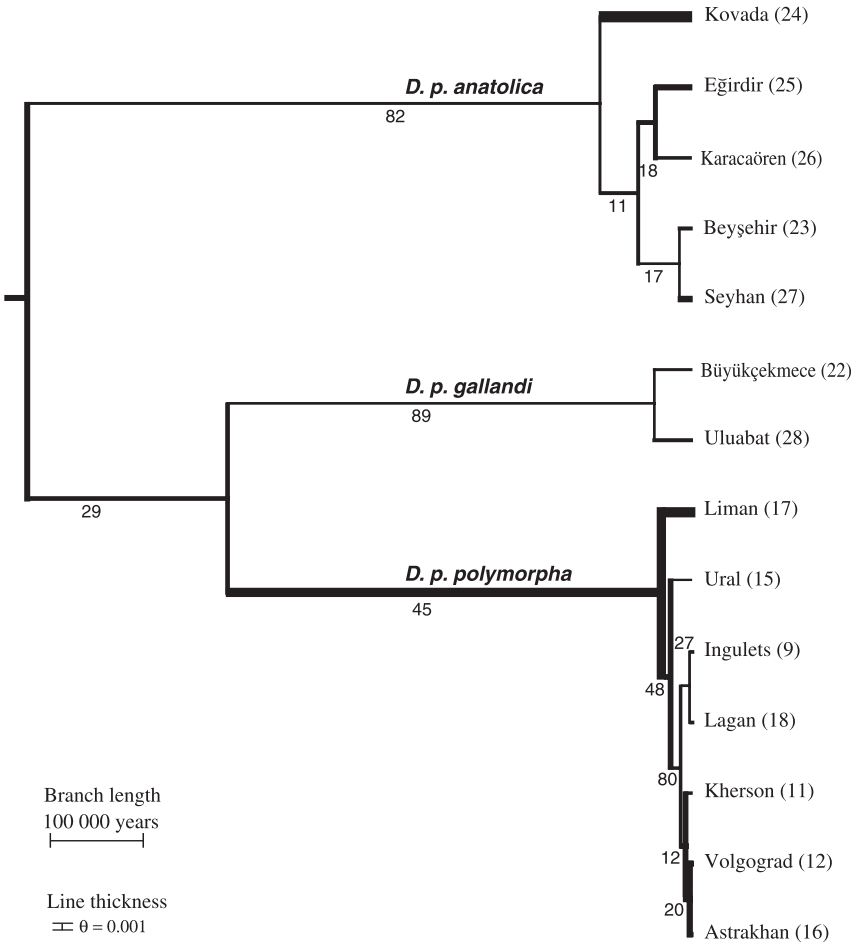

Fig. 3 Phylogeny of populations for the *Dreissena polymorpha* complex derived by coalescent analysis of mitochondrial cytochrome oxidase I (COI) gene sequences. The topology of the tree was inferred under a unique event polymorphism model of sequence evolution using the Bayesian program BATWING. Posterior probabilities are shown for each branch of the resulting consensus tree. Branch lengths and values of θ, reflecting population size, were estimated using the Bayesian program MCMCCOAL under an HKY model of sequence evolution, conditional on the topology obtained using BATWING. Branch width is proportional to the estimated value of θ for each population. The horizontal scaling bar reflects the molecular clock estimate of Marko (2002). For each population, the corresponding site numbers are specified.

mated value of T would correspond to a divergence time of 840 000 years between the populations in Lakes Prespa and Ohrid (Table 3). Under the additional assumption of one generation per year, estimates of θ would correspond to $N_e = 7.0 \times 10^6$ for Lake Prespa and $N_e = 3.1 \times 10^7$ for Lake Ohrid.

Discussion

We examined genetic imprints created by evolution in a region prone to disturbance. We hypothesized that such an evolutionary history might have produced distinctive genetic patterns at neutral markers and predisposed particular populations to invasiveness. Di Castri (1989) postulated that Old World species have a higher invasion potential than New World species due to a history of greater natural and anthropogenic disturbance. Some statistical support exists for such a bias in the directionality of invasion (Lonsdale 1999). The Ponto-Caspian basin has a history marked by fluctuations in environmental factors on multiple timescales (Svitoch *et al*. 2000; Reid & Orlova 2002).

The genetic patterns that we observed, including genetic signals of population bottlenecks and expansions, increased haplotype diversity in ancient stable lakes, and fragmentation into geographically disjunct, genetically distinct sets of populations, largely conformed to our expectations (see Introduction). Below, we retrace the evolutionary history of *Dreissena* in the Ponto-Caspian region. We then present a general synthesis examining the joint effects of habitat instability, passive aquatic dispersal, and high fecundity on the phylogenetic and coalescent genetic patterns that we observed in this genus.

Origins of the genus Dreissena

The origins of the genus *Dreissena*, like many other Ponto-Caspian fauna, trace back to the ancient marine Tethys Sea and subsequent brackish Paratethys Sea (Starobogatov 1994; Reid & Orlova 2002). Based on our analysis, the timing of the basal divergence event within *Dreissena* is far more ancient than previously recognized (68 Ma; 95% CI 35–125; also see Table S1). The basal split within *Dreissena* (Fig. 3) occurred during the Tethys Sea period and coincided

approximately with the Cretaceous–Tertiary mass extinction event (63 Ma) (Raup & Jablonski 1993). In addition, this estimated time range could implicate ecological changes, including periodic desiccation, which preceded final closure of the Tethys seaway connecting the Mediterranean and Indian Oceans. These ecological changes were apparently responsible for basal divergence (into eastern and western clades) in two other aquatic genera from this region, *Aphanius* fish (37.3 7 Ma) (Hrbek & Meyer 2003) and *Anguilla* eels (30–45 Ma) (Tsukamoto & Aoyama 1998).

Most previous authors have suggested a much more recent date for the basal split in this genus, based generally on morphological and fossil data, which can be ambiguous (Starobogatov 1994). Andrusov (1897) correctly believed the basal divergence between *Pontodreissena* and *Dreissena sensu stricto* to be an ancient Palaeogene event, although his rationale is suspect since he failed to recognize the monophyly of *Dreissena*. However, morphological analysis (reviewed in Nuttall 1990) potentially reconciles our older divergence dates with apparently more recent fossil dates. A key feature that distinguishes *Dreissena* from *Mytilopsis*, reduction of the apophysis, might actually be polyphyletic within *Dreissena* (Nuttall 1990). Thus, the morphologically identifiable forms of fossil *Dreissena* might have appeared independently in the lineages following the basal split within *Dreissena*, incorrectly suggesting recent derivations.

The only previous explicit estimate of the basal divergence based on molecular sequence data was 13.2 ± 2.2 Ma using the mitochondrial 16S gene (Stepien *et al*. 1999). However, the sole calibration point used here was divergence between the families Corbiculidae and Dreissenidae, anchored at the beginning of the Eocene (54 Ma). At the time of this analysis, it was widely believed that Dreissenidae was derived from the superfamily Corbiculidea (families Corbiculidae + Sphaeriidae) on approximately this timescale. However, recent molecular phylogenetic analysis (Park & ÓFoighil 2000) has clearly undermined this hypothesis, but is instead broadly consistent with Starobogatov's (1994) hypothesis that these three families (Corbiculidae, Sphaeriidae, Dreissenidae) last shared a common ancestor in the Ordovician (*c*. 500 Ma). Furthermore, allozyme data for *Dreissena bugensis* and *Dreissena polymorpha* (Nei's $D = 1.69$) (Spidle *et al*. 1994) also suggests a more ancient divergence date (> 32 Ma; allozyme clock of Grant 1987; see Maxson & Maxson 1979 regarding allozyme saturation).

Our phylogenetic analysis revealed four major clades within *Dreissena* (Fig. 2), rather than just the three subgenera currently recognized (*Pontodreissena*, *Carinodreissena*, and *Dreissena sensu stricto*) (Starobogatov 1994). Our tree topology for the three corresponding clades recapitulates recent findings of Therriault *et al*. (2004), also based on mtDNA sequence data. The additional clade, *Dreissena caputlacus*, is a recent discovery and its systematic placement had not been extensively considered (Fig. 2; Table 3). The morphological form of *D. caputlacus* closely resembles that of the fossil taxon *Dreissena diluvii* (Schütt 1993).

Two major divergence events within *Dreissena* (Fig. 2; *Carinodreissena*/*Dreissena sensu stricto* and *Pontodreissena*/*D. caputlacus*), one within each basal branch, both date to the early-mid Miocene and may reflect allopatric fragmentation events in the Paratethys Sea (Table 3). On this timescale, the Paratethys Sea began to separate from the Tethys near the Eocene–Oligocene boundary, 23 Ma. In the early-mid Miocene, seaway closures resulted in intermittent contact between the Eastern and Western Paratethys Seas, and between the Paratethys Sea, Mediterranean Sea, and Indo-Pacific Ocean (Rogl 1999; Harzhauser *et al*. 2002), potentially resulting in vicariant speciation. The divergence events in *Dreissena* might have coincided with a bloom in bivalve species diversity in the Paratethys, dating to the boundary between Karpatian and Badenian periods, 16 Ma (Harzhauser *et al*. 2003).

Ancient lakes in the Balkans

The genetic imprints left by divergence and expansion events within the major clades of *Dreissena* indicate that these events occurred much more recently than divergence among the major clades. Among the oldest divergence events within major clades are the colonizations of Lakes Ohrid and Prespa in the Balkans. These adjacent lakes appear to host ancient relict populations of *Dreissena stankovici*. Estimates of τ are consistent with mid-Pleistocene colonization events (Tables 2 and 3).

The high haplotype diversity and abundance of rare haplotypes (Fig. 1) conformed to our expectations, and suggest population growth and large effective population size in both lakes. Haplotype network and coalescent analyses suggest that colonization bottlenecks might have accompanied the invasion of both lakes (Fig. 1; Tables 1 and 2) and indicate that the Lake Ohrid population of *D. stankovici* might have been ancestral.

These ancient lakes, which formed *c*. 2–4 Ma (Stankoviç 1960; Watzin *et al*. 2002), would have provided relatively stable environments, in contrast to the fluctuations of the Ponto-Caspian Seas. In addition to their geological stability, their Balkan locations and the great depth of Lake Ohrid (maximum depth 289 m) should have insulated populations from disturbance during Pleistocene glacial cycles. These two lakes also harbour many other ancient fauna that have been geographically isolated for millions of years (Stankoviç 1960; Watzin *et al*. 2002). Endemism is extremely high among their benthic fauna (e.g. 86% of 50 known species of gastropods) (Watzin *et al*. 2002). In some cases, such as that of the sponge *Ochridospongia rotunda*, the closest related taxa inhabit distant Lake Baikal, another deep ancient tectonic lake (Watzin *et al*. 2002).

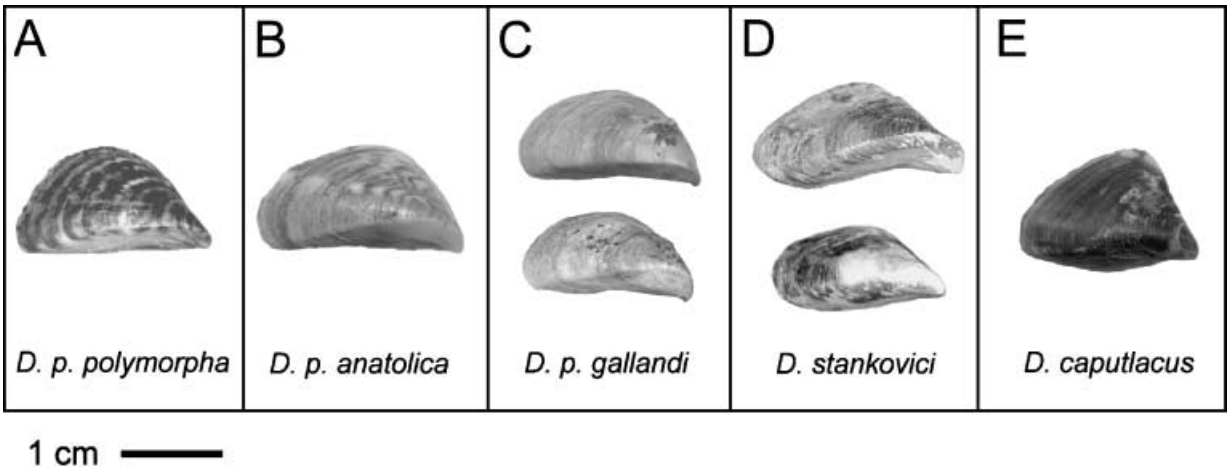

Fig. 4 Side view photographs of species of *Dreissena* (also see Appendix S2 in Supplementary materials). (A) *Dreissena polymorpha polymorpha* sampled from Lake Michigan (site 2); (B) member of the *D. polymorpha* complex sampled from Lake Kovada, Turkey (site 24), presumptive *D. polymorpha anatolica*; (C) member of the *D. polymorpha* complex sampled from Lakes Büyükçekmece (top) and Uluabat (bottom), Turkey (sites 22 and 28, respectively), presumptive *D. polymorpha gallandi*; (D) *Dreissena stankovici* sampled from Lakes Prespa (top) and Ohrid (bottom) (sites 20 and 21, respectively); (E) member of the genus *Dreissena* sampled from the Seyhan Dam (site 27) that morphologically resembled *Dreissena caputlacus*.

Evidence of an ancient demographic expansion event in Turkey

Dreissena caputlacus is a recently described species (Schütt 1993) that is highly distinctive, both morphologically (Fig. 4, Appendix S2 in Supplementary materials) and genetically (Fig. 2). Like *D. stankovici*, *D. caputlacus* showed a pattern of high haplotype diversity and possible evidence of an ancient expansion. The high haplotype diversity at Seyhan Dam reservoir was consistent with a large historical population size, conforming to our expectation for a stable lake habitat. The estimate of τ under a model of pure demographic expansion (Table 2) corresponded to an age of *c.* 920 000 years for the population, suggesting a mid-Pleistocene population expansion event (Table 3). However, the observed mismatch distribution, with an additional peak at 0 mismatches, fit a model of spatial expansion at time τ better than one of purely demographic expansion (yielding similar estimates of τ under both models) (Excoffier 2004). Such a pattern could also have been generated by a relatively recent postexpansion population bottleneck and might reflect invasion of the Seyhan Dam reservoir, from its endemic range within lakes of the nearby Gölbaşi tectonic basin (Schütt 1993).

Fragmentation of D. polymorpha

Our data were consistent with the hypothesis that geographic isolation of ancient lakes and seas in the Ponto-Caspian basin resulted in formation of a sibling species complex of *D. polymorpha* (Locard 1893; Kinzelbach 1992; Schütt 1993; Rosenberg & Ludyanskiy 1994; Starobogatov 1994). We found evidence of distinct evolutionary units within *D. polymorpha*, with no sharing of haplotypes between the Ponto-Caspian region and Turkey, or between northwestern and southcentral Turkey. Additional geographically disjunct subspecies of *D. polymorpha* have been described (Andrusov 1897; Kinzelbach 1992; Schütt 1993; Starobogatov 1994). Hrbek *et al*. (2004) noted an abundance of deeply diverged cryptic species in geologically complex areas of the Middle East, due to range fragmentation. All populations (and subspecies) within *D. polymorpha* exhibited low levels of haplotype diversity (especially relative to *D. caputlacus* and *D. stankovici*), and genetic signatures of recent population growth were detected. This pattern conformed to our expectation for fluctuating environments in the Ponto-Caspian Sea basin.

Of the *D. polymorpha* subspecies sampled, *D. polymorpha anatolica* (sites 23–27) was most genetically divergent (May *et al*. submitted; Figs 1–3), with a divergence estimate between *D. p. anatolica* and the other two subspecies (*D. p. polymorpha* + *D. p. gallandi*) of *c.* 730 000 BP (95% CI 290 000–1370 000 BP). The geographic location and timing of the split roughly coincided with a peak in species divergence events *c.* 2.23–1.07 Ma in two groups of endemic Mediterranean freshwater ichthyofauna, the genus *Chondrostoma* and the *Leuciscus* subgenus *Squalius* (Durand *et al*. 2003). Durand *et al*. (2003) hypothesized a wave of colonization in which Mesopotamian populations reached the Black and Caspian Seas through an extensive Anatolian inland lake system (Görür *et al*. 1995). Allopatric fragmentation resulted as the original Pliocene-era Anatolian lake was gradually eliminated with the advent of the Pleistocene due to tectonic uplift of the central Anatolian plateau (Görür *et al*. 1995; Westaway *et al*. 2004).

Dreissena polymorpha anatolica exhibited a genetic signature of an apparent population expansion *c.* 25 000 BP following a severe population bottleneck or founder event (Fig. 1; Table 1), a scenario congruent with geological data (Karabiyikoglu *et al.* 1999). Populations of *D. p. anatolica* currently inhabit Anatolian tectonic lakes in the Mediterranean zoogeographic province, predominantly within the Menderez-Taurus tectonic block (Kinzelbach 1992; Schütt 1993; Starobogatov 1994). Climatic cycles caused hydrological instability in this region during the late Pleistocene. High water levels last existed here during the last glacial maximum, 23 000–17 000 BP (Karabiyikoglu *et al.* 1999), coinciding with the population expansion event inferred from haplotype data. *D. p. anatolica* has not previously been reported to occur in the Seyhan Dam reservoir (site 27), and its presence here, like that of *D. caputlacus*, might have resulted from secondary introduction from its previously described range.

Dreissena polymorpha gallandi has been reported to occur in lakes bordering the Sea of Marmara south of Istanbul (Schütt 1993), within the Euro-Siberian zoogeographic province of Turkey. Its evolutionary history might be associated with the Marmara Basin, with its geological history of intermittent connection with the Black and Mediterranean Seas, and periodic existence as an isolated lake (Aksu *et al.* 1999). Contact between the Black and Marmara Seas was last restored *c.* 7000–10 000 BP (Aksu *et al.* 1999). Populations were monomorphic for haplotype E, unique to this subspecies, with coalescent analysis providing a divergence estimate of *c.* 510 000 years between *D. p. gallandi* and *D. p. polymorpha* (Table 3).

Split between D. polymorpha *populations in the Black and Caspian Seas*

The Black and Caspian Seas separated about 5 Ma, and have since experienced periods of interconnection and separation due to water level fluctuations ('transgressions' and 'regressions'). The connections between the seas during transgressions occurred along the low-lying Manych depression, the last such episode ending *c.* 9000 BP (Reid & Orlova 2002). These seas host a large number of endemic species adapted to brackish water. And due to their semi-independent history, these seas differ in endemic species composition, with 209 endemics restricted to the Caspian Sea and 20 restricted to the Black and Azov Seas (Mordukhai-Boltovskoi 1979).

Imperfect separation between these seas resulted in a region characterized by instability, with extirpations and recolonizations. Geological and fossil evidence suggest that *Dreissena* may have been extirpated in the Caspian Sea during a mass extinction event in the early Pliocene (*c.* 5.8 Ma) (Grigorovich *et al.* 2003). Subsequent recolonization of *D. polymorpha* and *D. rostriformis* appears to have occurred during the Akchagylian transgression (2–3.4 Ma), a period of temporary reconnection with the Black Sea (Grigorovich *et al.* 2003). More recently, during the Pleistocene, repeated incursions from the Mediterranean into the Black Sea resulted in periods of high salinity and extinction of many brackish-water Ponto-Caspian lineages within the Black Sea, with surviving lineages often confined to estuaries. Some Ponto-Caspian lineages may have subsequently recolonized the Black Sea from the Caspian, which experienced greater stability during this period. Such a recolonization, during the New Euxinian transgression (9000–19 000 BP) at the end of the Pleistocene, has been hypothesized in the case of *D. rostriformis*/*D. bugensis* (Babak 1983). Recent fluctuations in environmental conditions in the Black Sea may, in part, account for the strikingly small long-term effective population size of *D. polymorpha* in the Black Sea drainage (N_e approximately 80 000 given $\theta_{per\_site} = 0.00041$, assuming one generation/year).

Populations of *D. p. polymorpha* from Black and Caspian Sea drainages show evidence of a possible phylogeographic break, but relationships remain uncertain due to possible recent human-mediated introgression. Haplotypes C, D, and D2 were restricted to the Caspian Sea drainage, while A and B were geographically widespread. The presence of haplotypes A and B in areas proximal to the Volga River might have resulted from human activity (e.g. construction of the Don Canal). Samples from the Ural River (site 15), an independent Caspian Sea river system, contained only haplotypes C and D. The possibility of recent introgression of haplotypes A and B from the Black Sea into the Caspian Sea drainage via the Volga River is supported by both reports of other nonindigenous Black Sea-derived species (Grigorovich *et al.* 2003) and mtDNA haplotype data from other taxa (Cristescu *et al.* 2003). Coalescent analysis using BATWING showed apparent high support for one population subclade within *D. p. polymorpha* (Fig. 3). This subclade excluded two Caspian Sea drainage populations, including the Ural River. Because BATWING is limited to a nonreticulated representation of evolutionary history, it might not perfectly represent the secondary contact scenario above. However, the population phylogeny from BATWING is consistent with a split between Black and Caspian Sea drainages, but with blurring due to recent events (possibly human-mediated transport).

However, we cannot exclude the possibility that the presence of haplotypes A and B in the Caspian Sea pre-dated recent anthropogenic influences. This second hypothesis might be plausible because haplotypes A and B occur in the region of the final point of contact between the Black and Caspian Seas (i.e. low-lying Manych Depression/Northeastern Caspian). Under such a scenario, genetic differentiation between the Ural drainage and the Volga/Northeastern Caspian might reflect historical allopatric fragmentation, possibly due to an intervening salinity

barrier. However, the Ural delta is only 150 km from the Volga delta, and both rivers empty into the freshest (most Northern) part of the Caspian (salinity < 1 PSU in the surface layers; TES 1992, cited in Kaplin 1995), reducing the likelihood that haplotype differentiation would be long-standing and preserved on geological timescales.

To obtain a date of divergence between the Black and Caspian Seas under a hypothetical scenario in which haplotypes A and B were recent introductions, we performed coalescent analysis using the program IM for the Ural vs. Black Sea drainages (Kherson, site 11 + Ingulets, site 9). Here, the haplotype composition of the Ural sample was assumed to represent the pre-anthropogenic Caspian state. This analysis provided an estimate of *c.* 166 000 years (Table 3), potentially suggesting divergence following the ancient Euxinian transgression (350 000–250 000 BP). Furthermore, the timing of this split is roughly concordant with the timing of the inferred late-Pleistocene population expansion of *D. p. polymorpha*, *c.* 110 000 BP.

Phylogeographic breaks between the Black and Caspian Seas observed in *Dreissena* (e.g. *D. p. polymorpha* and *D. rostriformis*/*D. bugensis*) and dates for demographic events appear to be relatively shallow compared to those of other species. Studies based on mtDNA sequence data have found deep breaks across four species of amphipods and three cladocerans, with little population structure observed among sampling sites within each basin (Cristescu *et al*. 2003, 2004). Planktonic cladoceran species showed consistent mid-Pleistocene divergence times, with divergence dates *c.* 0.6–1.1 Ma (Cristescu *et al*. 2003). Benthic amphipod species exhibited older and more variable divergence dates, ranging from late Miocene (5.0–7.9 Ma) to early Pleistocene (1–1.6 Ma) (Cristescu *et al*. 2003, 2004). Overall, the species with more recent Black/Caspian divergence dates tend to show lower haplotype diversity. This trend might have arisen from greater extirpation/recolonization tendencies. Higher salinity tolerance and/or better capacity for upstream dispersal might have prevented extinctions in these cladoceran and amphipod species during salinity crises associated with marine transgressions. Additionally, long-lived populations of diapause eggs might have buffered against extinction in some microcrustaceans.

Origins of D. bugensis, *a second invasive Ponto-Caspian dreissenid*

A second invasive, the quagga mussel *D. bugensis*, is native to estuaries of the Black Sea (Marsden *et al*. 1996; Mills *et al*. 1996). This mussel is closely related to the Caspian Sea native *D. rostriformis*, with a divergence estimate of *c.* 500 000 BP (Table 3). This date roughly coincides with the appearance of specimens resembling *D. bugensis*, as well as the appearance of other Black Sea endemics, in the Black Sea basin fossil record (Starobogatov 1994). Starobogatov (1994) argued that *D. bugensis* was derived from the Black Sea endemic *D. tchaudae*. In contrast, Babak (1983) believed that *D. bugensis* arose from a recent, New Euxinian Epoch (9000–19 000 BP), reinvasion of the Black Sea by Caspian Sea *D. rostriformis*, following extinction of endemic Black Sea *Pontodreissena*. Based on data similar to our own, Therriault *et al*. (2004) argued that the low levels of genetic divergence in mtDNA sequences supported the hypothesis of Babak (1983). However, an estimate based on only three mutations in 606 nt (from one nonrecombining locus in reciprocally monophyletic taxa) is only moderately informative. Such data cannot fully resolve this issue because of the stochasticity in the coalescent (Edwards & Beerli 2000), but merely provide a loose upper bound on the divergence date. A molecular marker with higher resolution is required to accurately determine timing of this split.

A single haplotype was observed across two recognized subspecies of *D. rostriformis* that occupy habitats of differing depth. The differences between *D. rostriformis compressa* and *D. rostriformis distincta* in morphology and pigmentation exactly parallel differences between the deep and shallow forms of *D. bugensis* (Karpinsky 2005). Moreover, the morphological transition occurs at a similar depth in both species (Dermott & Munawar 1993; Karpinsky 2005). Deep and shallow forms of *D. bugensis* share a single identical mtDNA haplotype (Claxton *et al*. 1998; G. E. May & C. E. Lee, unpublished) and exhibit similarities in microsatellite allele frequencies (G. E. May & C. E. Lee, unpublished), possibly suggesting morphological plasticity rather than genetic differentiation. Lack of haplotype variation within both *D. rostriformis* and *D. bugensis* conforms to our expectations for dreissenids in the Ponto-Caspian Sea basin.

Synthesis: genetic imprints of a Ponto-Caspian heritage

The Ponto-Caspian Seas present a unique evolutionary environment. These seas experience major fluctuations in salinity, water level, temperature, and dissolved oxygen on both seasonal and geological timescales (Kaplin 1995; Svitoch *et al*. 2000). Although they contain unique species flocks that have radiated from a limited number of families, the total number of species is low (Remane & Schlieper 1971; Reid & Orlova 2002). Brackish-water habitats tend to be depauperate in species, and in any biogeographic region typically share a limited number of cosmopolitan taxa (Remane & Schlieper 1971). Furthermore, brackish-water habitats can constitute unsaturated biotopes in which species can achieve high population growth without displacing other fauna (Arndt 1989). Estuarine habitats in the Ponto-Caspian Seas thus provide a resource-rich but physically challenging and unstable environment.

The mode of dispersal of *Dreissena* could promote genetic fragmentation (Bohonak 1999). *Dreissena* possess pelagic, passively dispersed larvae. Poor dispersal and the

absence of mechanisms for dispersal counter to water flow contribute to the potential for allopatric fragmentation. Such fragmentation, occurring as a result of water level fluctuations (and perhaps salinity barriers), appears to have driven much of the observed genetic differentiation. Thus, taxa are confined to disjunct geographic regions (i.e. distinct seas and lake systems) (Fig. 2) (May *et al*. submitted).

Habitat instability and low dispersal ability would favour evolution of high reproductive potential and rapid maturation (Stearns 1992; Dillon 2000). For example, *D. p. polymorpha* gonadal tissue may constitute > 30% of body mass and a 15-mm female produces approximately 96 000 oocytes (Sprung 1991). Such high reproductive potential would facilitate escape to sites outside a zone of disturbance.

Frequent local extinctions and high reproductive potential might have been responsible for the genetic signals of bottlenecks and expansions that are common across *Dreissena* taxa and populations (Fig. 1; Tables 1 and 2). The genetic signals of such events would have been obscured if they were far in the past, or reduced if the number of colonizing propagules were large (i.e. weaker bottleneck). Environmental fluctuations outside the range of tolerance are likely to have been frequent in the Ponto-Caspian region, increasing the frequency of local extinctions. Moreover, with high reproductive potential and rapid maturation, once a few individuals colonized a site, numbers would have increased rapidly. Later migrants would have made little difference in the coalescent process and genetic diversity in the resulting population would have been low (with restoration of diversity requiring a mutational timescale). Colonizations at any given site would thus tend to be more recent and their genetic signatures would not have been erased by time.

The shape of the *Dreissena* phylogeny, with very long internal branches and a substantial number of branches near the tips, might also reflect a high extinction rate (Fig. 2). Increasing extinction rate relative to speciation rate increases internode length internally in a species tree, resulting in such a pattern (Nee *et al*. 1994). The dynamic instability of the Ponto-Caspian environment might have contributed to these features. Extinction proneness also increases with reduced dispersal power, due to a reduction in the ability to escape a zone of disturbance (Henle *et al*. 2004). As relict populations are confined to individual lakes or lake systems and most lakes are short-lived on a geological timescale, extinction of a lake would render the inhabiting populations of *Dreissena* extinct.

Adaptation to fresh water is polyphyletic (Fig. 2). Dreissenids are members of the bivalve subclass Heterodonta, from which frequent independent colonizations of fresh water have occurred (Deaton & Greenberg 1991). Heterodont mussels exhibit a capacity for hyperosmoregulation and hypercalcaemia (elevating haemolymph calcium concentrations) in dilute brackish water, traits that might serve as pre-adaptations for freshwater colonization (Deaton & Greenberg 1991). The fluctuating salinity of the Paratethys and Ponto-Caspian Seas and their estuaries, with periods of low salinity, might have facilitated adaptation to fresh water (Lee & Bell 1999). Theoretical models suggest that such temporal fluctuations facilitate evolutionary adaptation to an otherwise inaccessible environmental niche (Holt *et al*. 2004).

The Ponto-Caspian basin and possible antecedents of invasiveness

The Ponto-Caspian basin has been a source of many invasive species into the Great Lakes of North America (Lee & Bell 1999; Ricciardi & MacIsaac 2000). In particular, since 1985 fresh and brackish waters of the Ponto-Caspian region have contributed up to 70% of the invasive species into the Great Lakes (Ricciardi & MacIsaac 2000). Many Ponto-Caspian species, including zebra mussels (Hebert *et al*. 1989), quagga mussels (May & Marsden 1992), *Echinogammarus ischnus* (Witt *et al*. 1997), *Cercopagis pengoi* (MacIsaac *et al*. 1999), and *Bythotrephes cederstroemi* (Bur *et al*. 1986) were likely introduced via ballast water.

The fluctuating environmental conditions characteristic of the Ponto-Caspian Seas could promote the evolution of broad tolerance or the maintenance of genetic variance for tolerance upon which selection could act (Gillespie & Turelli 1989; Sasaki & Ellner 1997; Bürger & Gimelfarb 2002). Broad tolerance or high quantitative genetic diversity would increase the range of environments that a species could colonize. Broad tolerance might also facilitate survival during transport. For example, the relatively broad salinity tolerance characteristic of many Ponto-Caspian species might permit colonization of freshwater sites in North America despite partial mid-ocean ballast water exchange (Ricciardi & Rasmussen 1998). Furthermore, selection for high reproductive potential and rapid maturation could increase invasiveness on purely demographic grounds.

Populations of *Dreissena* that have evolved outside the dynamic environment of the Ponto-Caspian basin have not yet been observed to invade on a widespread scale. *D. stankovici*, *D. caputlacus*, *D. p. gallandi*, and *D. p. anatolica* are native to tectonic lakes. It is possible that the stable geological history and relative environmental constancy of tectonic lakes could have resulted in species less able to survive environmental shifts and, consequently, less able to readily invade. In addition to possible constraints due to physiology, life history, or adaptive potential, it is also possible that limited opportunity for transport is a factor, as shipping traffic from ancient Eurasian freshwater lakes is generally lower than traffic from Ponto-Caspian drainages.

The diversity of unique COI haplotypes in populations of *D. stankovici* and *D. caputlacus* (Fig. 1; Table 3 in May

et al. submitted) stands in sharp contrast to the depauperate set of haplotypes observed in dreissenid taxa such as *D. p. polymorpha* and *D. bugensis*, which are known to be highly invasive and that inhabit estuarine environments in Ponto-Caspian drainages. Ponto-Caspian species such as the invasive amphipod *Echinogammarus ischnus* (Cristescu *et al*. 2004) and several cladocerans (Cristescu *et al*. 2003) also exhibited low levels of intraspecific haplotype and nucleotide diversities at putatively neutral markers. Invasive populations of *E. ischnus* contained a single mitochondrial genotype of Black Sea origin (Cristescu *et al*. 2004). The low levels of genetic diversity of many Ponto Caspian species suggest that recent and severe bottlenecking events, possibly resulting from environmental fluctuations, not only occurred in populations of dreissenids but also in other Ponto-Caspian species.

The Ponto-Caspian Seas present an environment that might foster evolution of traits associated with invasiveness. Evolution in such an environment might leave unique phylogenetic and population genetic signatures, in part through selection of particular life history characteristics. However, patterns vary across taxa, with microcrustaceans showing much deeper phylogenetic splits between Black and Caspian Sea populations than that seen in dreissenids. Relatively little research has been done on evolutionary genetics of invasive species (Lee 2002). Characterization of additional taxonomic groups from this region might permit identification of cohesive patterns across species and yield general insights.

Acknowledgements

Funding was provided by Wisconsin Sea Grant Institute, R/LR-87-PD and R/LR-88, and NSF DEB-0130543 to C. E. Lee. The UW Anna Birge Undergraduate Research Fund contributed funds for Michael Gerner's collection of samples from Russia, Macedonia and Greece. Director Alexander Alimov and Vadim Panov sponsored Michael Gerner as a visiting scholar at the Zoological Institute of the Russian Academy of Sciences. Greg Gelembiuk was supported by Sea Grant R/LR-87-PD, NSF Predoctoral Fellowship, and UW Anna Birge Fund. Gemma May was supported by Sea Grant R/LR-88 and UW Bunde and Anna Birge Funds. The following assisted with sample collections: Ebru Unal, Yuriy Slynko, Alexander Gorbunov, F. de Valence, Dawn Dittman, Helena Belyaeva, Shafia Powell, T. Naeher, F. Shakirova, Christoph Shubart, Fabian Skibinski, J. de Leeuw, Konstantin, and Galina. Dianna Padilla provided contacts with scientists in Russia. Emily Ruell aided in sequencing. Bill Feeny produced photographs (Fig. 4). Louis Bernatchez, Jenny Boughman, Heather Free, Suzanne Peyer, and Marijan Posavi provided helpful suggestions and editorial comments.

Supplementary materials

The supplementary materials are available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2816/MEC2816sm.htm

Appendix S1 Molecular clock analyses

Appendix S2 Physical descriptions of newly collected samples of *Dreissena*

Table S1 Alternative calibrations and inferred divergence time between *D. polymorpha* and *D. bugensis*. Estimates of proportional difference based on Kimura 2-parameter model (K2P) are presented for the specified contrasts and genes. The mean rate for each gene is used to estimate an age for the *D. polymorpha*/*D. bugensis* split. The set of species chosen for analysis of COI is largely identical to that used by Luttikhuizen *et al*. (2003).

References

Aksu AE, Hiscott RN, Yasar D (1999) Oscillating Quaternary water levels of the Marmara Sea and vigorous outflow into the Aegean Sea from the Marmara Sea–Black Sea drainage corridor. *Marine Geology*, **153**, 275–302.

Andrusov NI (1897) Fossile und lebende Dreissenidae Eurasiens. *Trudy S-Peterbugskago Obshchestva Estestvoispitatalei (Department of Geology and Mineralogy)*, **25**, 1–683.

Arndt EA (1989) Ecological, physiological and historical aspects of brackish water fauna distribution. In: *Reproduction, Genetics and Distributions of Marine Organisms: Proceedings of the 23rd European Marine Biology Symposium, Swansea, 5–9 September 1988* (eds Ryland JS, Tyler PA), pp. 327–338. Olsen and Olsen, Denmark.

Babak EV (1983) The Pliocene and Quaternary Dreissenidae of the Evsinsk Basin. *Trudy Paleontologiceskogo Instituta Akademii Nauk SSSR*, **204**, 1–104.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, **74**, 21–45.

Bur MT, Klarer DM, Krieger KA (1986) First records of a European cladoceran, *Bythotrephes cederstroemi*, in Lakes Erie and Huron. *Journal of Great Lakes Research*, **12**, 144–146.

Bürger R, Gimelfarb A (2002) Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genetical Research*, **80**, 31–46.

Claxton WT, Wilson AB, Mackie GL, Boulding EG (1998) A genetic and morphological comparison of shallow- and deep-water populations of the introduced dreissenid bivalve *Dreissena bugensis*. *Canadian Journal of Zoology*, **76**, 1269–1276.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657.

Coan EV, Scott PV, Bernard FR (2000) *Bivalve Seashells of Western North America: Marine Bivalve Mollusks from Baja California to Arctic Alaska*. The Santa Barbara Museum of Natural History, Santa Barbara.

Cox LR, Newell ND, Boyd DW *et al*. (1969) Bivalvia. In: *Treatise on Invertebrate Paleontology, Part N Mollusca 6, vol.1&2* (ed. Moore RC). Geological Society of America, Boulder, Colorado.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Cristescu ME, Hebert PDN, Onciu TM (2003) Phylogeography of Ponto-Caspian crustaceans: a benthic-planktonic comparison. *Molecular Ecology*, **12**, 985–996.

Cristescu MEA, Witt JDS, Grigorovich IA, Hebert PDN, MacIsaac HJ (2004) Dispersal of the Ponto-Caspian amphipod *Echinogammarus ischnus*: invasion waves from the Pleistocene to the present. *Heredity*, **92**, 197–203.

Deaton LE, Greenberg MJ (1991) The adaptation of bivalve molluscs to oligohaline and fresh waters: phylogenetic and physiological aspects. *Malacological Review*, **24**, 1–18.

Dermott R, Munawar M (1993) Invasion of Lake Erie offshore sediments by *Dreissena*, and its ecological implications. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 2298–2304.

di Castri F (1989) History of biological invasions with special emphasis on the old world. *Biological Invasions: a Global Perspective* (eds J.A. Drake, H.A. Mooney, F. di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek & M. Williamson), pp. 1–30. John Wiley & Sons, New York.

Dillon RT (2000) *The Ecology of Freshwater Molluscs*. Cambridge University Press, Cambridge, UK.

Durand J-D, Bianco PG, Laroche J, Gilles A (2003) Insight into the origin of endemic Mediterranean ichthyofauna: phylogeography of *Chondrostoma* genus (Teleostei, Cyprinidae). *Journal of Heredity*, **94**, 315–328.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 727–727.

Felsenstein J (1998) *PHYLIP: Phylogeny Inference Package (version 3.6)*. Department of Genetics, University of Washington, Seattle, Washington.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Geary DH, Magyar I, Muller P (2000) Ancient Lake Pannon and its endemic molluscan fauna (Central Europe; Mio-Pliocene). *Advances in Ecological Research*, **31**, 463–482.

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics*, **121**, 129–138.

Görür N, Sakınç M, Barka A, Akkök R, Ersoy S (1995) Miocene to Pliocene palaeogeographic evolution of Turkey and its surroundings. *Journal of Human Evolution*, **28**, 309–324.

Grant WS (1987) Genetic divergence between congeneric Atlantic and Pacific ocean fishes. In: *Population Genetics and Fishery Management* (eds Ryman N, Utter F), pp. 225–246. University of Washington Press, Seattle, Washington.

Grassly NC, Rambaut A (2003) *TREEVOLVE version 1.3: A program to simulate the evolution of DNA sequences under different population dynamic scenarios*. Oxford University, Oxford.

Grigorovich IA, Therriault TW, MacIsaac HJ (2003) History of aquatic invertebrate invasions in the Caspian Sea. *Biological Invasions*, **5**, 103–115.

Harzhauser M, Mandic O, Zuschin M (2003) Changes in Paratethyan marine molluscs at the Early/Middle Miocene transition: diversity, palaeogeography and palaeoclimate. *Acta Geologica Polonica*, **53**, 323–339.

Harzhauser M, Piller WE, Steininger FF (2002) Circum-Mediterranean Oligo-Miocene biogeographic evolution — the gastropods' point of view. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **183**, 103–133.

Hebert PDN, Muncaster BW, Mackie GL (1989) Ecological and genetic studies on *Dreissena polymorpha* (Pallas): a new mollusc in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 1587–1591.

Henle K, Davies KF, Kleyer M, Margules C, Settele J (2004) Predictors of species sensitivity to fragmentation. *Biodiversity and Conservation*, **13**, 207–251.

Hey J, Nielsen R (2005) *IM — A program for fitting the Isolation w/ Migration model using a Markov chain Monte Carlo method*. Rutgers University. http://lifesci.rutgers.edu/~heylab/ProgramsandData/Programs/IM/IMdocumentation.pdf

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. *Computational and Applied Biosciences*, **8**, 189–191.

Holt RD, Barfield M, Gomulkiewicz R (2004) Temporal variation can facilitate niche evolution in harsh sink environments. *American Naturalist*, **164**, 187–200.

Hrbek T, Meyer A (2003) Closing of the Tethys Sea and the phylogeny of Eurasian killifishes (Cyprinodontiformes: Cyprinodontidae). *Journal of Evolutionary Biology*, **16**, 17–36.

Hrbek T, Stolting KN, Bardakci F, Kucuk F, Wildekamp RH, Meyer A (2004) Plate tectonics and biogeographical patterns of the *Pseudophoxinus* (Pisces: Cypriniformes) species complex of central Anatolia, Turkey. *Molecular Phylogenetics and Evolution*, **32**, 297–308.

Huelsenbeck JP, Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, **53**, 904–913.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Kaplin PA (1995) The Caspian: its past, present, and future. In: *In Enclosed Seas and Large Lakes of Eastern Europe and Middle Asia* (ed. Mandych AF), pp. 71–118. SPB Academic Publishing, Amsterdam, The Netherlands.

Karabiyikoglu M, Kuzucuoglu C, Fontugne M, Kaiser B, Mouralis D (1999) Facies and depositional sequences of the Late Pleistocene Gocu shoreline system, Konya basin, Central Anatolia: implications for reconstructing lake-level changes. *Quaternary Science Reviews*, **18**, 593–609.

Karpinsky MG (2005) *Dreissena* (*Pontodreissena*) *rostriformis*. Caspian Fisheries Research Institute, Astrakhan, Russia (www.caspianenvironment.org/biodb/eng/zoobenthos/Dreissena%20 (Pontodreissena)%20rostriformis/main.htm).

Kelley CT (1998) *MATLAB/FORTRAN software for Iterative Methods for Optimization*. www4.ncsu.edu/eos/users/c/ctkelley/www/matlab_darts.html.

Kinzelbach R (1992) The main features of the phylogeny and dispersal of the zebra mussel *Dreissena polymorpha*. In: *The Zebra Mussel Dreissena Polymorpha* (eds Neumann D, Jenner HA), pp. 5–17. Gustav Fischer, Stuttgart.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

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.

Locard A (1893) Les Dreissensia du système européen d'après la collection Bourguignat. *Revue Suisse Zoologie*, **1**, 113–185.

Lonsdale WM (1999) Global patterns of plant invasions and the concept of invasibility. *Ecology*, **80**, 1522–1536.

Luttikhuizen PC, Drent J, Baker AJ (2003) Disjunct distribution of highly diverged mitochondrial lineage clade and population subdivision in a marine bivalve with pelagic larval dispersal. *Molecular Ecology*, **12**, 2215–2229.

MacIsaac HJ, Grigorovich IA, Hoyle JA, Yan ND, Panov VE (1999)

Invasion of Lake Ontario by the Ponto-Caspian predatory cladoceran *Cercopagis pengoi*. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1–5.

Marko PB (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution*, **19**, 2005–2021.

Marsden JE, Spidle AP, May B (1996) Review of genetic studies of *Dreissena* spp. *American Zoologist*, **36**, 259–270.

Maxson LR, Maxson RD (1979) Comparative albumin and biochemical evolution in plethodontid salamanders. *Evolution*, **33**, 1057–1062.

May GE, Gelembiuk GW, Panov VE, Orlova MI, Lee CE (submitted) Molecular ecology of zebra mussel invasions. *Molecular Ecology*. doi: 10.1111/j.1365-294X.2005.02814.x

May B, Marsden JE (1992) Genetic identification and implications of another invasive species of dreissenid mussel in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 1501–1506.

Mills EL, Rosenberg G, Spidle AP *et al*. (1996) A review of the biology and ecology of the quagga mussel (*Dreissena bugensis*), a second species of freshwater dreissenid introduced to North America. *American Zoologist*, **36**, 271–286.

Moore RC (1969) *Treatise on Invertebrate Paleontology, Part N Mollusca 6, vol.1&2*. Geological Society of America, Boulder, Colorado.

Mordukhai-Boltovskoi PD (1979) Composition and distribution of Caspian fauna in the light of modern data. *Internationale Revue der Gesamten Hydrobiologie*, **64**, 1–38.

Nee S, Holmes EC, May RM, Harvey PH (1994) Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **344**, 77–82.

Nuttall CP (1990) Review of the Caenozoic heterodont bivalve superfamily Dreissenacea. *Palaeontology*, **33**, 707–737.

Park JK, Ó Foighil D (2000) Sphaeriid and corbiculid clams represent separate heterodont bivalve radiations into freshwater environments. *Molecular Phylogenetics and Evolution*, **14**, 75–88.

Pohlo R (1982) Evolution of the Tellinacea (Bivalvia). *Journal of Molluscan Studies*, **28**, 245–256.

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

Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics*, **18**, 559–563.

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.

Raup DM, Jablonski D (1993) Geography of end-Cretaceous marine bivalve extinctions. *Science*, **260**, 971–973.

Reid D, Orlova MI (2002) Geological and evolutionary underpinnings for the success of Ponto-Caspian species invasions in the Baltic Sea and North American Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1144–1158.

Remane A, Schlieper C (1971) *Biology of Brackish Water*. 2nd edn. John Wiley & Sons, New York.

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

Ricciardi A, Rasmussen JB (1998) Predicting the identity and impact of future biological invaders: a priority for aquatic resource management. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1759–1765.

Rogl F (1999) Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geologica Carpathica*, **50**, 339–349.

Rosenberg G, Ludyanskiy ML (1994) A nomenclatural review of *Dreissena* (Bivalvia: Dreissenidae), with identification of the quagga mussel as *Dreissena bugensis*. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1474–1484.

Rozas J, Rozas R (1999) DNASP (version 3): an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Sasaki A, Ellner S (1997) Quantitative genetic variance maintained by fluctuating selection with overlapping generations: variance components and covariances. *Evolution*, **51**, 682–696.

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

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: A software for population genetics data analysis. Version 2.000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Schütt H (1993) Die Gattung *Dreissena* im Quartär Anatoliens (Bivalvia: Eulamellibranchiata: Dreissenacea). *Archiv Molluskenk*, **122**, 323–333.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Spidle AP, Marsden JE, May B (1994) Identification of the Great Lakes quagga mussel as *Dreissena bugensis* from the Dnieper River, Ukraine, on the basis of allozyme variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1485–1489.

Sprung M (1991) Costs of reproduction: a study on metabolic requirements of the gonads and fecundity of the bivalve *Dreissena polymorpha*. *Malacologia*, **33**, 63–70.

Stankoviç S (1960) *The Balkan Lake Ohrid and Its Living World*. Monographiae Biologicae, Den Haag.

Starobogatov YI (1994) Taxonomy and paleontology. In: *Freshwater Zebra Mussel: Dreissena Polymorpha (Pall.) (Bivalvia, Dreissenidae): Systematics, Ecology, Practical Meaning* (ed. Starobogatov YI), pp. 18–46. Russian Academy of Sciences (Nauka Press), Moscow.

Stearns SC (1992) *The Evolution of Life Histories*. Oxford University Press, Oxford, UK.

Stepien CA, Hubers AN, Skidmore JL (1999) Diagnostic genetic markers and evolutionary relationships among invasive dreissenoid and corbiculoid bivalves in North America: phylogenetic signal from mitochondrial 16S rDNA. *Molecular Phylogenetics and Evolution*, **13**, 31–49.

Strayer D (1999) Invasion of fresh waters by saltwater animals. *Trends in Ecology &Evolution*, **14**, 448–449.

Svitoch AA, Selivanov AO, Yanina TA (2000) The Pont-Caspian and Mediterranean basins in the Pleistocene (paleogeography and correlation). *Oceanology*, **40**, 868–881.

Swofford DL (1998) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods) Version* 4.0. Sinauer Associates, Sunderland, Massachusetts.

Therriault TW, Docker MF, Orlova MI, Heath DD, MacIsaac HJ (2004) Molecular resolution of the family Dreissenidae (Mollusca: Bivalvia) with emphasis on Ponto-Caspian species, including

first report of *Mytilopsis leucophaeata* in the Black Sea basin. *Molecular Phylogenetics and Evolution*, **30**, 479–489.

Tsukamoto K, Aoyama J (1998) Evolution of freshwater eels of the genus *Anguilla*: a probable scenario. *Environmental Biology of Fishes*, **52**, 139–148.

Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics*, **145**, 847–855.

Watzin MC, Puka V, Naumoski TB (2002) *Lake Ohrid and Its Watershed, State of the Environment Report*. Lake Ohrid Conservation Project, Tirana, Albania, and Ohrid, Macedonia.

Westaway R, Pringle M, Yurtmen S *et al.* (2004) Pliocene and Quaternary regional uplift in western Turkey: the Gediz River terrace staircase and the volcanism at Kula. *Tectonophysics*, **391**, 121–169.

Wilson I, Balding D, Weale M (2003) BATWING *User Guide (version 1.02)*. Department of Mathematical Sciences, University of Aberdeen. www.maths.abdn.ac.uk/~ijw/downloads/batwing/batguide.pdf.

Witt JDS, Hebert PDN, Morton WB (1997) *Echinogammarus ischnus*: another crustacean invader in the Laurentian Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 264–268.

Supplementary Materials for
Gelembiuk, GW, GE May, and CE Lee. 2006. Phylogeography and systematics of zebra mussels and related species. Molecular Ecology. 15:1021–1031.

Appendix I: Molecular Clock Analyses

To further corroborate the approximate validity of dates estimated using the clock of Marko (2002) and COI sequence data, we performed alternative estimations of the age of the basal divergence within *Dreissena*. We employed alternative calibration points within the bivalve fossil record and sequence data from mitochondrial COI and 16S and nuclear 18S genes (see below). Specifically, a set of three to four fossil-based divergence times (Table 4, below) were used to calculate a mean molecular clock rate for each gene, based on proportional difference between pairs of sequences estimated using a Kimura 2-parameter model of sequence evolution. For each of the three genes, this rate was then used to estimate the age of the *Dreissena* basal split using sequence data from *D. polymorpha* and *D. bugensis*.

To provide a loose lower bound on the date of the basal *Dreissena* divergence event, we anchored the divergence between the *Mytilopsis* + *Dreissena* clade *versus Congeria* at the lower Eocene boundary (54 mya). This clock analysis was performed using MrBayes 3.01, under a birth-death prior, using the codon-structured GTR+G model. We note that the fossil record for the period preceding the Eocene is sparse (Nuttall 1990) and the actual basal divergence date for Dreissenidae might have been much earlier than 54 mya. This analysis assumes the validity of the phylogeny of Starabogatov (1994). Starabogatov concluded that *Congeria* descended from the fossil taxon *Eocongeria*, with a split from *Mytilopsis+Dreissena* "earlier than the Eocene" due to the diversity and abundance of both lineages in the later Eocene. Analysis of COI data, using MrBayes 3.01 and outgroups *Corbicula flumina* and *Macoma bathica* (GenBank accession# U47647, AY162256), conforms to this phylogeny and confirms basal divergence of the *Congeria* lineage (*Mytilopsis+Dreissena* clade posterior probability estimated as 0.99). This analysis resulted in an estimate of 26 mya (95%CI 14-42) for the minimum age of the basal divergence event within *Dreissena*.

Table 4. Alternative calibrations and inferred divergence time between *D. polymorpha* and *D. bugensis*. Estimates of proportional difference based on Kimura 2-parameter model (K2P) are presented for the specified contrasts and genes. The mean rate for each gene is used to estimate an age for the *D. polymorpha/D. bugensis* split. The set of species chosen for analysis of COI is largely identical to that used by Luttikhuizen *et al.* (2003).

Contrast	Divergence Time (Reference)	Gene
Macoma vs. *Donax*	90-140 mya (Pohlo 1982)	COI
		18S
Sinonovacula constricta vs. *Donax/Macoma*	206-251 mya (Pohlo 1982)	COI
Mytilus edulis vs. *Mytilus californianus*	30 mya (Coan *et al.* 2000)	COI
		16S
		18S
Mytilus vs. *Adamussium/Pecten*	400-480 mya (Moore 1969)	16S
Barbatia lima vs. *Arca ventricosa*	144-206 mya (Cox *et al.* 1969)	COI
		18S
Chamelea gallina vs. *Venus verrucosa*	25-40 mya (Cox *et al.* 1969)	16S

	Inferred Divergence Time	
D. polymorpha vs. *D. bugensis*	89 mya estimate	COI
	41 mya estimate	16S
	92 mya estimate	18S

GenBank Sequences used in Molecular Clock Calibration Analyses

16S rRNA: AJ243882 (*Adamussium colbecki*), AJ548762 (*Chamelea gallina*), AJ548763 (*Venus verrucosa*), AY650084 (*Pecten maximus*), AF023600 (*Mytilus californianus*), U68770 (*Mytilus californianus*), AY350784 (*Mytilus edulis*), U22885 (*Mytilus galloprovincialis*), AF507049 (*Dreissena polymorpha*), AF507047 (*Dreissena bugensis*)

COI: AB076935 (*Arca ventricosa*), AB076931 (*Barbatia lima*), AB040842 (*Donax cuneatus*), AB076949 (*Sinonovacula constricta*), AF443221 (*Macoma balthica*), AB040844 (*Donax faba*), U68776 (*Mytilus californianus*), AF241963 (*Mytilus edulis*)

18S rRNA: AY570554 (*Macoma balthica*), AJ309018 (*Donax trunculus*), X90960 (*Arca noae*), AF207646 (*Barbatia barbata*), X91974 (*Barbatia virescens*), L33448 (*Mytilus edulis*), L33449 (*Mytilus californianus*), AF305702 (*Dreissena polymorpha*), AF305703 (*Dreissena bugensis*).

Appendix II: Physical Descriptions of Newly Collected Samples of *Dreissena*

Dreissena polymorpha anatolica (Fig. 4B; sites 23-27)
Shells were relatively small, rather elongated, and almost rectangular, with equally-sized valves that were moderately inflated. The umbo bent slightly, but often acutely (in a beaked fashion), and was close to the ventral field. The byssal hole is closer in proximity to the posterior than to the anterior. The central portion of the dorsal rim was often relatively straight and parallel to the ventral rim. A wing-like ridge near the ventral field on the valves extended from the umbo and tapered off near to the anterior of the ventral field. The shells were greyish/cream in color and often had a faint dark striped pattern similar to that of *D. p. polymorpha*.

Dreissena polymorpha gallandi (Fig. 4C; sites 22, 28)
Shells were moderately large and relatively elongated, with equally-sized valves that were moderately inflated. The umbo bent deeply and was close to the ventral field. The byssal hole was closer in proximity to the posterior than to the anterior. The posterior portion of the shell was slightly bulbous from a side view, with the dorsal rim forming a long sweeping arc. A wing-like ridge near the ventral field on the valves extended from the umbo and tapered off to the midpoint of the ventral field. The shells were white, tan, greenish-tan, and light blue with no distinguishable patterns.

Dreissena caputlacus (Fig. 4E; site 27)
The type specimens occur in lakes near Gölbaşi, Turkey, but have not previously been observed in the Seyhan Dam reservoir near Adana, Turkey (site 27). Shells were moderate/small in size and triangular, with equally-sized valves that were inflated. The byssal hole was very small and narrow and closer in proximity to the posterior than to the anterior. A wing-like ridge near the ventral field on the valves extended from the umbo and tapered off to the midpoint of the ventral field. The dorsal field was very wide; the dorsal edge was characterized by a pointed "wing" that was narrow and nearly flat. From a ventral view, the ventral shell margin was slightly

sinusoidal. The exterior of the shell was very dark brown, nearly black in color. The interior of the shell was blue and black.