*Journal of Heredity* 2011:102(2):196–206
doi:10.1093/jhered/esq120
Advance Access publication January 27, 2011



# Homogeneity at Nuclear Microsatellite Loci Masks Mitochondrial Haplotype Diversity in the Endangered Fanshell Pearlymussel (*Cyprogenia stegaria*)

J. Paul Grobler, Jess W. Jones, Nathan A. Johnson, Richard J. Neves, and Eric M. Hallerman

From the Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0321(Grobler, Jones, Johnson, Neves, and Hallerman); the Department of Genetics, University of the Free State, Bloemfontein, South Africa (Grobler); and the U.S. Fish and Wildlife Service, Blacksburg, VA (Jones).

Address correspondence to J. Paul Grobler at the address above, e-mail: groblerjp@ufs.ac.za.

## Abstract

We report on multiple patterns of differentiation and connectivity in the fanshell pearlymussel (*Cyprogenia stegaria*), based on different markers. Knowledge of genetic variation and genetic connectivity among remaining populations of this federally endangered species is needed to initiate implementation of the species recovery plan. We collected tissue samples from 96 specimens from the Green, Rolling Fork, and Licking Rivers, tributaries to the Ohio River, and the Clinch River, a tributary to the Tennessee River, providing broad coverage of the current distributional range of the species. Results from 7 nuclear DNA microsatellite markers suggested minimal population-level differentiation, whereas a mitochondrial DNA (mtDNA) marker (*ND1*) exhibited significant differentiation between *C. stegaria* in the Clinch River and the Ohio River populations. The *ND1* data also confirm the existence of 2 distinct mtDNA lineages in the genus that transcends species boundaries. Further analyses suggest that the disproportionally strong signal from 2 very divergent *ND1* lineages possibly masks finer-grained structure in the Ohio River population, based on one of the mtDNA lineages only. We recommend further sampling to confirm the absence of one lineage from the upper Clinch River drainage and suggest that provisional management guidelines should limit reciprocal exchanges among *C. stegaria* populations from the Clinch River and those in the Ohio River system.

**Key words:** Cyprogenia stegaria, *endangered species, genetic structure, microsatellites, mtDNA, Unionidae*

The fanshell pearlymussel *Cyprogenia stegaria* (Rafinesque, 1820) (Bivalvia: Unionidae) is endemic to the Ohio, Tennessee, and Cumberland Rivers of the Mississippi River drainage, USA. (Jones and Neves 2002). The species reaches a maximum size of approximately 80 mm and has a shell that is subcircular in outline (Johnson 1980). The fanshell pearlymussel was listed as endangered by the U.S. Fish and Wildlife Service (USFWS) in 1990 (USFWS 1990) and was placed in the critically endangered category of the IUCN Red List (Groombridge 1994). The eggs of adult female mussels are fertilized in the gills, which serve as brooding chambers, by sperm taken in while filtering. The larvae (glochidia) of most freshwater mussel species are parasitic and must attach to the gills or body of specific fish or amphibian hosts to transform into juveniles and hence disperse to new habitats. Gene flow among freshwater mussel populations is thus primarily driven by dispersal of fish hosts (Elderkin et al. 2007).

The 760 km of river habitat currently inhabited by *C. stegaria* represents less than 10% of its former range (Jones and Neves 2002). The primary causes of decline of *C. stegaria* are impoundments, navigation projects, pollution, and dredging (USFWS 1991). Reproducing and viable populations are known from only 4 rivers: the upper Clinch River, a headwater tributary of the Tennessee River drainage in Tennessee (TN) and Virginia, and the Green, Licking, and Rolling Fork Rivers of the Ohio River drainage in Kentucky (KY) (USFWS 1991). However, these remaining populations are isolated from each other, with contemporary gene flow severely restricted or prevented by numerous dams (USFWS 1991). A very small but reproducing population exists in the middle reach of the Tennessee River in Alabama and TN (Parmalee and Bogan 1998), and other populations of questionable viability have been reported from the Muskingum River, Ohio; Wabash River, Illinois; East Fork White River and Tippecanoe River, Indiana;

Kanawha River, West Virginia; Tygarts Creek, KY; and Cumberland River, TN (USFWS 1990). The long life span of freshwater mussels (Haag and Rypel, forthcoming) raises uncertainty over the viability of these populations because observation of live or fresh-dead individuals is no guarantee that the population is still reproducing. The long life span also adds a measure of uncertainty to estimates of genetic diversity. Gaggiotti and Vetter (1999) showed that generation overlap can buffer the effect of environmental fluctuations on the level of genetic variability maintained by a population. However, it is also possible that the inclusion of very old individuals in a sample can result in a distorted view of current levels of diversity in recruiting individuals.

Recovery strategies for *C. stegaria* include augmentation of extant populations and reestablishment of populations into historical habitat where water quality has been sufficiently improved and appropriate fish hosts are present (USFWS 1991). Therefore, the management implications of genetic structuring among populations must be considered and has been discussed by Mulvey et al. (1997), Machordom et al. (2003), Mock et al. (2004), Grobler et al. (2006), Mock et al. (2010) and other authors for numerous mussel species. For example, several published studies have demonstrated significant genetic structure among populations of *Potamilus inflatus* (Roe and Lydeard 1998), *Lampsilis hydiana* (Turner et al. 2000), *Lexingtonia dolabelloides* (Grobler et al. 2006), *Amblema plicata* (Elderkin et al. 2007), and *Fusconaia flava* (Burdick and White 2007). Evidence of such genetic structure suggests that translocations of individuals between distantly related populations should be discouraged in order to maintain evolutionarily significant units (ESUs) (Ryder 1986; Waples 1991, 1995; Moritz 1994, 2002; Crandall et al. 2000).

Notwithstanding known examples of significant population genetic structure, it is also possible that little or no genetic structuring occurs across large geographic areas of a species' distributional range. For example, minimal genetic structuring was reported for sampled demes of *Quadrula quadrula* occurring over hundreds of river miles of the Mississippi and Ohio River systems (Berg et al. 1998). Because gene flow among mussel populations is likely driven primarily by host dispersal, these authors ascribed the lack of structuring to the mobility of its hosts, which include several catfish species. Roe and Lydeard (1998) suggested that incorrect application of the ESU concept could hinder rather than aid in the recognition and conservation of invertebrate biodiversity. Erroneous assignment of populations as ESU's could promote gene pool fragmentation if management plans involve limitations on inducing artificial gene flow. Such an outcome would be contradictory to the aims of conserving connectivity.

More than a decade after the first discussion of the ESU concept by Waples (1991), Moritz (1994) and others, the challenge remains to accurately define populations and test the power of genetic markers to detect real population boundaries and describe connectivity. Waples and Gaggiotti (2006) reviewed commonly used population definitions and highlighted the theoretical and empirical problems of detecting reproductive cohesion. Similarly, Palsbøll et al. (2006) reviewed the identification of management units in species and found that conclusions about population status can change, even when using the same data, as different kinds of population criteria and genetic analyses are implemented.

The purposes of this study were to assess the genetic structure among 4 reproducing populations of *C. stegaria* and determine levels of genetic diversity within each population. Such data and analyses can aid in identifying the most appropriate source and recipient populations for management and recovery of the species.

## Materials and Methods

### Sample Collection and DNA Extraction

Individuals of *C. stegaria* were sampled from the following 4 localities across the range of the species (Figure 1): The Green River, KY ($n = 30$), Rolling Fork River, KY ($n = 21$), and Licking River, KY ($n = 20$) of the Ohio River drainage; and the Clinch River, TN ($n = 25$) of the Tennessee River drainage. Samples from the first 3 localities were each collected in local areas over limited stream distances (200–300 m) and these presumably represent populations with high reproductive cohesion, whereas the sample from the Clinch River included 5 sites spread over 33 km stream distance (Swan Island [river mark 277.1 km, $n = 9$], Sneedville/Fall Branch [287.5 km, $n = 5$], Frost Ford [291.7 km, $n = 6$], Brooks Island [294.9 km, $n = 4$], and Wallen Bend [309.6 km, $n = 1$]). Approximately 20–50 mg of mantle tissue was collected nonlethally from each mussel sampled. Tissue was preserved in 95% ethanol prior to DNA isolation. Total DNA was isolated using the Purgene DNA extraction kit. The success and quality of DNA extraction was verified using 0.8% agarose gels. The concentration of DNA extracts was quantified by fluorescence assay.

### Genetic Analysis

Differentiation among and within sampling locations was determined using both nuclear DNA microsatellites and mitochondrial DNA (mtDNA) markers. Because these 2 categories of markers are subject to different modes of inheritance, and often reflect differentiation on different evolutionary time spans, a combination of data from both types of molecular markers was considered the best strategy to ensure a high degree of confidence in the results.

### Amplification of DNA Microsatellites and Statistical Analyses

Samples were genotyped at 8 dinucleotide microsatellite loci. We used cross-species amplification of primers developed for the oystermussel (*Epioblasma capsaeformis*) by Jones et al. (2004). All polymerase chain reactions (PCRs) mixtures

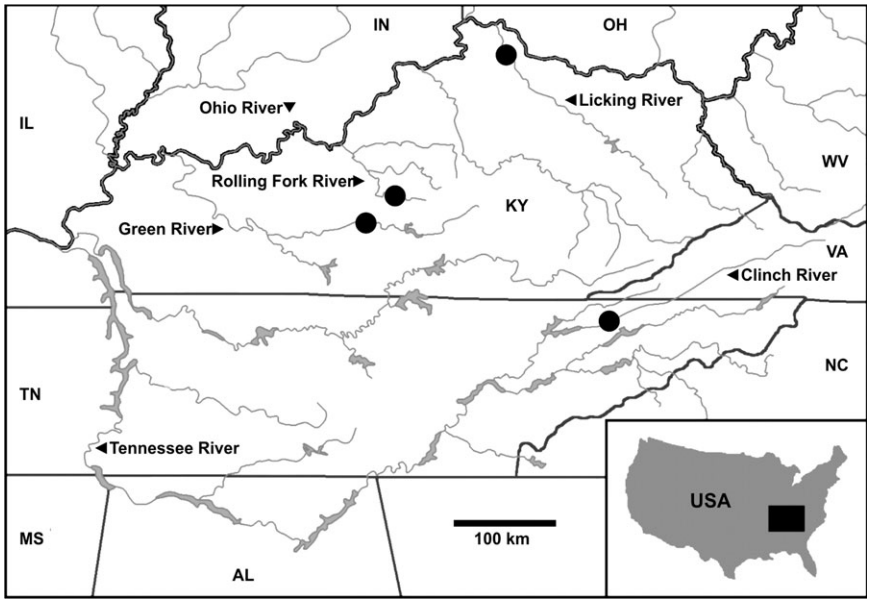


**Figure 1.** The geographic range of *Cyprogenia stegaria* populations sampled in the Ohio and Tennessee River systems. The symbol (●) indicates sampling sites.

consisted of 100 ng genomic DNA, 1× PCR buffer, 2.5 mM $MgCl_2$, 0.4 mM dNTPs, 1.5 *U* Ampli*Taq* Gold DNA polymerase, and 1 μM of each primer, with $ddH_2O$ added to a total volume of 20 μl. Forward primers were labeled with the fluorescent primers 6-FAM or HEX. The thermal cycling profile consisted of an initial 95 °C for 7 min; followed by 35 cycles each of 94 °C for 40 s, 58 °C for 40 s, and 72 °C for 60 s; with a final extension step of 5 min at 72 °C and a final hold at 4 °C. All amplification reactions were conducted separately for each locus, but differently labeled PCR products were pooled for fragment analyses. Microsatellite lengths were resolved on an Applied Biosystems (ABI) 3100 automated sequencer and scored using ABI proprietary software.

Statistical analyses of microsatellite data were conducted to 1) assess the suitability of the markers developed by Jones et al. (2004) for population genetic analysis of *C. stegaria*, 2) estimate genetic divergence among populations, and 3) estimate within-population genetic diversity for each population. We used ARLEQUIN 3.0 (Excoffier et al. 2005) to screen for genotypic disequilibrium among the 8 loci and to test for deviations from expected Hardy–Weinberg equilibrium, using a Markov chain with 1000 steps. The presence of null alleles was assessed using MICROCHECKER (Van Oosterhout et al. 2004).

Differentiation between population pairs was quantified using $F_{ST}$ (Wright 1965), with no assumption of a stepwise mutation model. We calculated $F_{ST}$ in ARLEQUIN software, using 10 000 permutations. The ARLEQUIN software also was used to conduct an analysis of molecular variance (AMOVA) to determine the hierarchical distribution of total within- and among-population genetic diversity. To test for isolation by distance, a Mantel test (Smouse and Long 1992) was used to measure the correlation between $F_{ST}$ values and geographic stream distance using PASSAGE, version 1 (Rosenberg 2003).

STRUCTURE software (Pritchard et al. 2000) was used to infer population structure and assign individuals to populations. An admixture ancestry model with correlated allele frequencies was used to determine population structure of *C. stegaria* individuals. The true number of populations (*K*) was estimated by calculating the parameter $-\ln\Pr(X|K)$ for $K = 1$–8 and with 5 independent runs for each *K*. All runs consisted of a burn-in of 100 000 steps, followed by 200 000 iterations.

Genetic diversity within populations was quantified using average expected heterozygosity ($H_e$ – Nei, 1978) and average number of alleles per locus, using POPGENE software (Yeh et al. 1999). We also calculated allelic richness ($R_s$), which takes into account the variation in sample size, using FSTAT (Goudet 2001).

We used a sequential Bonferroni correction to account for Type I errors associated with all multiple pairwise comparisons (Rice 1989).

### mtDNA Sequencing and Statistical Analyses

The first subunit of NADH dehydrogenase (*ND1*) was sequenced to determine mtDNA-based differentiation between populations. We used primers and PCR amplification conditions reported in Buhay et al. (2002) and Serb et al. 2003). Primer sequences were forward, 5′-TGGCAGAAAAGTGCATCAGATTAAAGC-3′ and reverse, 5′-CCTGCTTGGAAGGCAAGT GTACT-3′. The PCR reaction mixture consisted of 100 ng of genomic DNA, 1× PCR buffer, 4.0 mM $MgCl_2$, 0.4 mM dNTPs, 1.0 μM each

primer, and 1.5 *U* Ampli*Taq* Gold DNA polymerase, with $ddH_2O$ added to a total volume of 25 µl. The thermal cycling profile consisted of an initial 95 °C for 8 min; followed by 35 cycles of: 94 °C for 40 s, 50 °C for 60 s, and 72 °C for 90 s; with a final extension step at 72 °C for 2 min; and a final hold at 4 °C.

In preparation for sequencing, PCR products were purified using a Qiagen DNA Purification kit to remove any remaining primers. PCR products were sequenced with ABI Big Dye 3.1. The reaction mixture consisted of 30 ng of PCR product, 5 µM primer, 2 µl Big Dye 3.1 (diluted 1:1), with $ddH_2O$ added for a final volume of 7.5 µl. The cycle reaction consisted of 30 cycles, each of 94 °C for 30 s, 50 °C for 15 s, and 60 °C for 4 min. The products were cleaned by centrifugation through hydrated Sephadex (Sigma) in Millipore filter plates, dried down, and resuspended in Hi-Di Formamide (ABI). The samples were denatured at 95 °C for 5 min and cooled to 4 °C for 2 min before being loaded on the ABI DNA Analyzer 3730 for processing. SEQUENCHER software (version 4.11) was used to align and edit sequences.

The goals of the mtDNA analyses were similar to those for nuclear microsatellite data to 1) determine the scale of genetic structuring among populations, 2) describe within-population diversity for each population, and 3) compare diversity values calculated for each population. We used ARLEQUIN to calculate AMOVA values to estimate the hierarchical distribution of mtDNA genetic differentiation within and among populations, using $\Phi_{ST}$ (from the absolute number of nucleotide differences) and with 16 000 permutations. Pairwise population differentiation values ($F_{ST}$) were quantified using uncorrected *p*-distance (number of nucleotide substitutions), which is the mean number of changes per site since 2 taxa last shared a common ancestor and corrected average pairwise distance. Genetic diversity within populations was quantified as the number of polymorphic sites, nucleotide diversity ($\pi$), haplotype diversity ($h$), and the mean number of pairwise differences, as implemented in ARLEQUIN software.

A phylogenetic analysis was conducted using Bayesian inference in MrBayes v3.0b4 (Huelsenbeck and Ronquist 2001). MrBayes was run for 2 million generations, sampling trees every 100 generations, and posterior probabilities were computed after a burn-in of 20 000 generations. Modeltest 3.7 (Posada and Crandall 1998) was used to choose the most likely model of evolution for the *Cyprogenia* mtDNA data set. We used 6 haplotypes of the congeneric *C. aberti* to serve as a reference group to gauge the relative magnitude of genetic differentiation values estimated among populations of *C. stegaria*. These haplotypes occurred in 15 individuals collected from the St Francis River, Missouri and Verdigris River, Kansas (Gen Bank accession numbers EF125918 to EF125922, and EF125879, with the latter identical to EF125880 in *C. stegaria*). In addition, *ND1* haplotypes of various conspecific lampsiline mussel species were obtained from GenBank and used as out-group taxa to include: *Dromus dromas* (Jones WJ, unpublished data), *Medionidus conradicus* (AY158746), *Lemiox rimosus* (AY655104), *Potamilus alatus* (AY655119), *Lampsilis cardium* (EF446096), *Lampsilis ovata* (AY613797), *Lampsilis fasciola* (DQ220721), *E. brevidens* (AY094378), *E. capsaeformis* (DQ208591), *Villosa iris* (DQ445185), *Ligumia recta* (EF213055), *Lampsilis higginsii* (EF213061), *Actinonaias ligamentina* (AY655085), *Lampsilis siliquoidea* (UAUC882), and *Amblema plicata* (UAUC147). A minimum-spanning haplotype network of *Cyprogenia* haplotypes was constructed using a median joining approach (Bandelt et al. 1999) in the program NETWORK, version 4.5.

## Results

### Amplification of Microsatellites, Population Differentiation, and Genetic Diversity

Microsatellites were successfully amplified in *C. stegaria* using 8 of 10 primer pairs developed for *E. capsaeformis*. Size ranges of microsatellites and numbers of alleles per locus for all *C. stegaria* populations are presented in Table 1. The sizes of microsatellites generated during the current study are comparable with size ranges reported by Jones et al. (2004), with overlapping size ranges at 7 loci and range borders

**Table 1** Allele size ranges, number of alleles per locus (in brackets), average expected heterozygosity ($H_e$), average number of alleles per locus ($A$), and allelic richness ($R_s$) in 4 *Cyprogenia stegaria* populations

| Locus | Green | Rolling Fork | Licking | Clinch |
|---|---|---|---|---|
| *Ecap1* | 146–154 (4) | 146–150 (3) | 146–154 (3) | 146–148 (2) |
| *Ecap2* | 129–151* (12) | 133–155* (11) | 123–151 (12) | 127–151* (14) |
| *Ecap4* | 127–157 (13) | 131–157 (11) | 125–149 (11) | 131–151 (12) |
| *Ecap5* | 177–233 (19) | 179–213 (12) | 177–221 (13) | 177–219* (16) |
| *Ecap6* | 233–249 (7) | 233–245 (4) | 233–247 (6) | 233–247 (7) |
| *Ecap7* | 101–147* (22) | 99–145 (17) | 103–145 (13) | 99–143 (18) |
| *Ecap8* | 131–153* (12) | 131–155 (12) | 135–153 (6) | 127–155 (13) |
| *Ecap10* | 94–110 (7) | 96–110 (6) | 102–110 (5) | 98–110 (6) |
| $H_e$ | 0.751 (±0.084) | 0.712 (±0.087) | 0.732 (±0.089) | 0.775 (±0.075) |
| $A$ | 12.0 (±6.63) | 9.29 (±5.09) | 8.14 (±4.10) | 10.71 (±5.68) |
| $R_s$ | 7.373 (±3.583) | 6.442 (±3.248) | 7.131 (±3.749) | 7.205 (±3.481) |

Polymorphism is 100% in all cases. *Indicates loci with significant ($P < 0.05$) deviations from Hardy–Weinberg equilibrium of genotypes.

differing by only 2 repeat units at locus *Ecap 2*. Tests for genotypic disequilibrium between microsatellite loci within populations indicated that all loci were in genotypic equilibrium ($\chi^2$ test, $P < 0.05$, with a sequential Bonferroni correction). Significant deviations ($P < 0.05$) from expected Hardy–Weinberg equilibrium of genotypes were observed at 3 loci (Table 1). Isolated deviations restricted to single populations were regarded as indicative of population-specific processes, whereas the value in the Clinch population is possibly indicative of a Wahlund effect. For locus *Ecap2*, significant deviations due to heterozygote deficiencies were found in 75% of populations. Results from MICROCHECKER suggested that null alleles may be present at this locus only, and results for *Ecap2* were therefore not included in subsequent analysis.

Results from AMOVA indicated that 98.43% of overall variation in *Cyprogenia* is found within populations and only 1.57% among populations. There were no $F_{ST}$ values (Table 2) between any population pair that indicated significant ($P < 0.05$) differentiation (with and without application of the Bonferroni correction). Results of the Mantel test provided no support for a model of isolation by distance in *C. stegaria*, with no significant correlation between absolute geographic distance and $F_{ST}$ values ($P = 0.246$).

The assignment test analysis conducted in STRUCTURE showed the highest likelihood at a population structure of $K = 1$. This $K$ value yielded the highest average −lnPr values and the lowest deviation among 5 independent runs (with −lnPr for $K = 1$–8 equal to −2142.12 ± 8.88; −2146.86 ± 14.08; −2203.58 ± 68.56; −2181.84 ± 47.62; −2468.28 ± 419.40; −2504.98 ± 317.27; −2444.78 ± 213.87, and −2655.92 ± 579.88). Distributions based on all $K$ values greater than one were associated with a high degree of symmetrical admixture among clusters for all individuals.

Average expected heterozygosity was similar in all 4 populations of *C. stegaria*, with values ranging from 0.71 to 0.77 (Table 1). Average number of alleles per locus ranged from 8.1 in the Licking population and 9.3 in the Rolling Fork to 10.7 in the Green population and 12.0 in the Clinch. Trends in allelic richness among populations also were similar (Table 1).

## mtDNA Diversity and Phylogeny

An 888-bp fragment of the *ND1* gene was obtained for analysis in all sampled individuals after sequences were aligned and edited. A total of 38 *ND1* haplotypes could be identified among 96 individuals of *C. stegaria* (GenBank accession numbers EF125880 to EF125917).

Only 4 haplotypes were shared between 2 or more populations, with unique haplotypes in all populations (Table 3). The most common haplotype was found in *C. stegaria* from the Green, Licking, and Rolling Fork Rivers, but this haplotype was absent from populations from the Clinch River that did not share any haplotypes with any other population. The *C. stegaria* haplotypes could be divided into 2 distinct lineages, separated by 137 mutation events. Lineage one (L1) contained individuals from all rivers sampled and included 40 polymorphic sites, whereas lineage 2 (L2) contained individuals from the Green, Licking, and Rolling Fork Rivers but not from the Clinch River. This lineage was variable at 10 nucleotide sites. Both lineages were also found in the *C. aberti* haplotype sequences used as reference group.

Results from MODELTEST showed that variation in the mtDNA haplotypes was best described by the Hasegawa–Kishino–Yano plus Gamma model. A phylogram based on this model of substitution and the Bayesian tree evaluation criterion is presented in Figure 2. The phylogram confirms a pattern of 2 divergent mtDNA lineages but without structure corresponding to population boundaries, except for the exclusion of lineage L1 from the Clinch River population. This clear separation between the 2 lineages is fully consistent with clusters observed in the minimum-spanning haplotype network (Figure 3). There were no patterns in the network that reflect population boundaries; however, there is a cluster of closely related haplotypes from the Clinch River, 1–2 mutational steps removed from the main cluster of *C. stegaria* haplotypes.

**Table 2** Population differentiation among 4 populations of *Cyprogenia stegaria*

| | | i. | ii. | | iii. | |
|---|---|---|---|---|---|---|
| Population pair | | $F_{ST}$: microsatellites | *p*-distance: all sequences | Corrected pairwise difference: all sequences | *p*-distance: Lineage I | Corrected pairwise difference: Lineage I |
| Green | -Rolling | 0.0 | 0.088 | 5.027 | 0.559* | 3.855* |
| | -Licking | 0.002 | 0.009 | 0.394 | 0.671* | 5.236* |
| Rolling | -Licking | 0.008 | 0.098 | 4.531 | 0.001 | 0.0 |
| Clinch | -Green | 0.0 | 0.292* | 12.461* | 0.379* | 2.294* |
| | -Rolling | 0.0 | 0.545* | 34.273* | 0.210* | 1.281* |
| | -Licking | 0.0 | 0.283* | 11.147* | 0.390* | 2.639* |

(i) $F_{ST}$ values from microsatellite data; (ii) the number of nucleotide substitutions among populations expressed as uncorrected *p*-distance and corrected pairwise differences, based on all *ND1* haplotypes; and (iii) uncorrected *p*-distance and corrected pairwise differences from haplotypes of *ND1*-Lineage 1 only. Values marked with asterisk indicate significant ($P < 0.05$) differences between population pairs after Bonferroni correction.

**Table 3** The mtDNA haplotypes, GenBank accession numbers, haplotype frequencies, and coefficients of genetic diversity for four populations of *Cyprogenia stegaria*

| | | *C. stegaria* | | | |
|---|---|---|---|---|---|
| Haplotype | GenBank no. | Green River | Rolling Fork River | Licking River | Clinch River |
| Csteg01 | EF125880 | 0.107 | 0.050 | 0.105 | — |
| Csteg02 | EF125881 | — | 0.250 | 0.316 | — |
| Csteg03 | EF125882 | 0.036 | 0.100 | 0.263 | — |
| Csteg04 | EF125883 | 0.250 | 0.150 | — | — |
| Csteg05 | EF125884 | 0.286 | — | — | — |
| Csteg06 to Csteg13 | EF125885 to EF125892 | 0.036 each | — | — | — |
| Csteg14 | EF125893 | — | 0.150 | — | — |
| Csteg15 | EF125984 | — | 0.150 | — | — |
| Csteg16 to Csteg19 | EF125985 to EF125898 | — | 0.050 each | — | — |
| Csteg20 | EF125899 | — | — | 0.105 | — |
| Csteg21 | EF125900 | — | — | 0.105 | — |
| Csteg22 to Csteg23 | EF125901 to EF125902 | — | — | 0.053 each | — |
| Csteg24 | EF125903 | — | — | — | 0.150 |
| Csteg25 | EF125904 | — | — | — | 0.150 |
| Csteg26 | EF125905 | — | — | — | 0.100 |
| Csteg27 to Csteg38 | EF125906 to EF125917 | — | — | — | 0.050 each |
| Polymorphic sites (*n*) | | 123 | 122 | 119 | 22 |
| Nucleotide diversity ($\pi$) | | 0.060 ± 0.030 | 0.066 ± 0.033 | 0.059± 0.030 | 0.006 ± 0.003 |
| Haplotype diversity (*h*) | | 0.464 | 0.450 | 0.368 | 0.750 |
| Number of pairwise differences | | 53.423 ± 23.812 | 58.315 ± 26.299 | 53.064 ± 24.018 | 4.916 ± 2.499 |

Calculation of AMOVA values showed that 78.86% of total mitochondrial haplotype variation occurred within the 4 *C. stegaria* populations, whereas 21.14% occurred among populations. Genetic divergence values among populations, based on the number of nucleotide substitutions among populations and expressed as uncorrected *p*-distance and corrected pairwise difference values, are presented in Table 2. Pairwise comparisons between populations indicated significant ($P < 0.05$) differentiation between the Clinch River population and populations from the Green, Rolling Fork, and Licking Rivers. Differentiation between any pairwise populations in the Ohio River was not significantly different from zero. To test for the influence of the 2 highly divergent lineages on differentiation values, AMOVA was repeated with lineages included as an extra hierarchical level. Results showed that differences between L1 and L2 accounted for 95.31% of overall variation, with 1.98% contributed by differences between rivers and 2.71% found within river populations. To compensate for this disproportionate effect of L2, AMOVA was also applied using data from L1 only. This analysis revealed a hierarchical distribution with 57.07% of total variation within populations and 42.93% of variation attributed to the 4 rivers. Genetic divergence values among population pairs, based on uncorrected *p*-distance and corrected K2P values from L1 only, indicated significant ($P < 0.05$) differentiation between all population pairs except those from the Rolling Fork and Licking Rivers (Table 2).

Several genetic diversity measures indicated that within-population diversity was closely comparable for the Green, Rolling Fork, and Licking River populations but markedly lower in the Clinch River population: nucleotide diversity ranged from 0.059 to 0.066 compared with 0.006; number of polymorphic sites ranged from 119 to 123 compared with 22; and mean number of pairwise differences ranged from 53.064 to 58.315 compared with 4.916. However, haplotype diversity was higher in the Clinch River population (0.750) compared with the other 3 populations (0.368–0.464), indicating the lower level of genetic diversity in the Clinch River population was due to the absence of Lineage 1.

### Assessment of mtDNA Lineages Based on Analysis of DNA Microsatellites

Following the finding of 2 divergent mtDNA lineages of *C. stegaria* which overlap population and species boundaries, microsatellite-based coefficients of population differentiation in this species were recalculated to test for support of these lineages based on analysis of nuclear DNA markers. All individuals in each population were divided into 2 groups based on possession of either lineage L1 or L2 (except the Clinch River where only L2 was found). No differentiation was observed between these 2 groups, with $F_{ST} = 0.0$ ($P = 0.644$).

## Discussion

### Population Genetic Structure

Analysis of nuclear microsatellites suggests that little to no genetic structuring exists among populations of *C. stegaria*, notwithstanding the spatial distance (up to 1800 km stream distance) and anthropogenic barriers that now separate remaining populations of the species following the construction of dams and locks from the 1930s to 1960s. Because *C. stegaria* is no longer found in 90% of its former

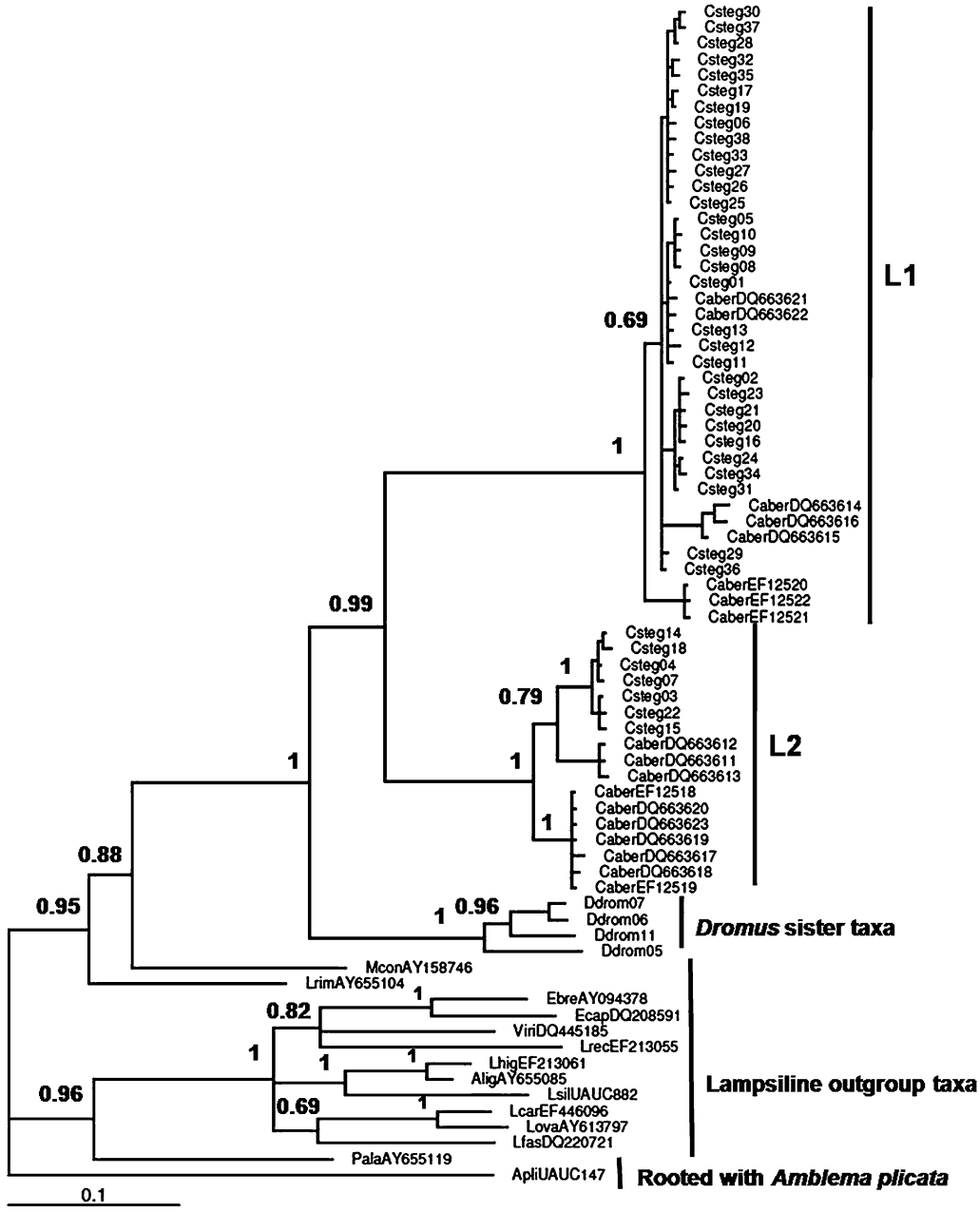


**Figure 2.** A 50% majority-rule consensus tree (phylogram) showing phylogenetic relationships among haplotypes of *Cyprogenia stegaria*, *C. aberti*, *Dromus dromas*, lampsiline out-group taxa, and *Amblema plicata*, which was used to root the tree. The 2 *Cyprogenia* lineages (L1 and L2) are shown, which include haplotypes from both species of *Cyprogenia*. The tree is based on *ND1* mtDNA sequences and was evaluated using Bayesian inference (3 000 000 generations, burn-in = 30 000 generations). Numbers above the branches are the posterior credibility values indicating proportion of trees containing the inferred nodes. Labels correspond to haplotype numbers used in Table 3.

distribution area, it is not known whether this nuclear DNA homogeneity reflects current connectivity or rather the signature of former gene flow, still maintained in large remnant populations. A widely used frequency-based measure of differentiation, $F_{ST}$, demonstrated that all pairwise comparisons of *C. stegaria* populations were not significantly greater than zero, and the AMOVA results indicated that most variation was found within populations (98.43%), with very little between populations (1.57%). This lack of genetic structure among populations of *C. stegaria* was also confirmed using an individual-based Bayesian assignment test; a method that shows high power to discriminate an individual's population of origin in some published studies (Manel et al. 2002, Orsini et al. 2004). In

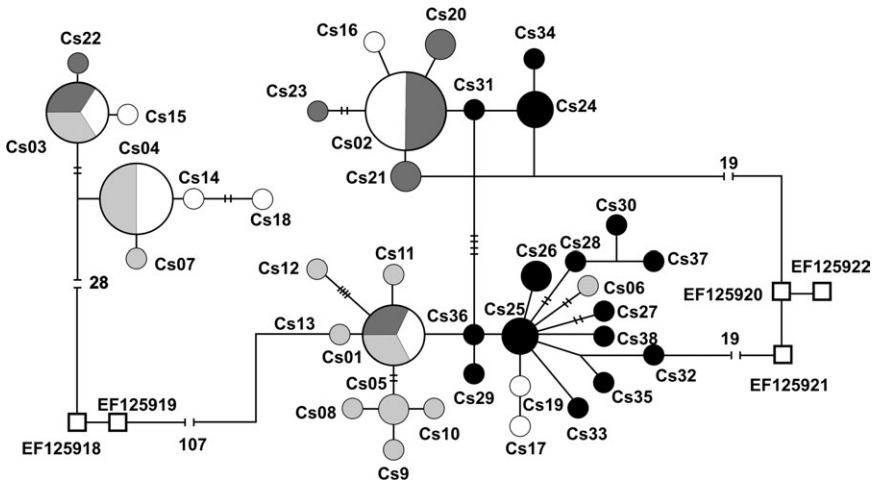


**Figure 3.** Minimum-spanning network between haplotypes of *Cyprogenia stegaria*. Nodes represent haplotypes, with the size of each node representing the number of individuals that share that haplotype. Numbers used to denote haplotypes follow Table 3. Cross-bars reflect the number of mutational events between specific haplotype pairs; a clear line connecting 2 circles represents a single event. For $n > 4$ mutational events, numbers next to interrupted lines are used to indicate the number of mutations. Colors used to indicate population(s) of origin of haplotypes: Green ●, Rolling Fork ○, Licking ●, and Clinch ●; and with *C. aberti* haplotypes □ for comparison.

its application to *C. stegaria*, all individuals from all populations were grouped into a single population.

In contrast to the nuclear DNA data, the mtDNA data showed greater structure within and among populations of *C. stegaria*. The haplotypes of *C. stegaria* could be divided into 2 distinct lineages, lineage L1 that occurred in individuals from all rivers sampled, and lineage L2 that is seemingly less abundant, which occurred in individuals collected from the Green, Licking, and Rolling Fork Rivers but not in individuals collected from the Clinch River. Both of these mtDNA lineages transcend species boundaries and thus include the haplotypes of *C. aberti* (Figure 2).

The finding of 2 lineages in *Cyprogenia* corroborates those of Serb (2006) and Serb and Barnhart (2008) who reported the existence of 2 divergent lineages in their studies of *C. aberti* from the western Mississippi River basin. The hypothesis that one of the mtDNA lineages is the result of inheritance infidelity of a male mitotype (see Zouros et al. 1992 and Stewart et al. 1995) has been tested in *C. aberti* and found to be false (Serb J, personal communication). Therefore, the 2 mtDNA lineages have most likely descended maternally in the genus *Cyprogenia* either through incomplete lineage sorting or hybridization. The accumulation of species-specific mutations suggests that the event(s) leading to this haplotype distribution is relatively ancient and not due to recent interspecific gene flow. Ellsworth et al. (1994) noted that species defined by morphological characteristics may appear polyphyletic in terms of mtDNA lineages for many generations following a speciation event. Another possibility is that one of the lineages was introgressed into *Cyprogenia* from another species. However, we used a variety of out-group taxa to address this possibility and as shown in Figure 2, all *Cyprogenia* haplotypes are monophyletic.

The absence of lineage L2 from the Clinch river sample is regarded as particularly significant because L2 was consistently found in all local populations sampled in the Ohio River system, yet it was not found in a sample spanning 5 localities over 33 km of stream distance in the Clinch River. Two scenarios possibly explain this absence of L2. First, the Clinch River population was part of the larger *C. stegaria* population during the development of the 2 mtDNA lineages but experienced a significant bottleneck (in situ) thereafter, leading to loss of L2. Second, the Clinch River population could have been isolated from conspecific populations before development of the 2 lineages. In this scenario, the lower genetic diversity in the Clinch River population could reflect a founder event during colonization of the Tennessee River basin by *C. stegaria*. However, this scenario seems unlikely because allele frequency divergence at nuclear loci between the Clinch River population and those in the Ohio River should be correspondingly high, which they were not. Furthermore, although mtDNA differentiation between the Clinch River population and those in the Green, Rolling Fork, and Licking Rivers were significantly different from zero in all cases, and the AMOVA results indicated that 78.86% of total variation resided within populations and 21.14% between populations, much of this differentiation is being driven by the absence of L2 in the Clinch River population. Indeed, an AMOVA with lineages included as an extra hierarchical level suggests that 95.31% of total variation is due to differences between L1 and L2, compared with only 1.98% ascribed to populations with the same lineage in separate rivers.

A reappraisal of differentiation in *C. stegaria* based on L1 reveals possible fine-grained *ND1*-based structure that was swamped by the strong signal of 2 divergent mtDNA

lineages. The data from L1 suggest differentiation among most sampled locations, except between the Rolling Fork and Licking Rivers (2 adjacent tributaries to the Ohio River) which were panmictic. This pattern could signal a model of migration broadly based on the stepping-stone migration model, where the geographically most distant populations show significant differentiation but where some intervening populations might show uninterrupted gene flow. This structure suggested by L1 data was not found in the nuclear microsatellite. However, we note that data from mtDNA may reveal differentiation not evident from nuclear DNA because the nature of mtDNA markers theoretically result in a 4-fold reduction in effective population size ($N_e$) compared with nuclear loci (Birky et al. 1983), with an associated increase in rates of drift.

### Population Genetic Diversity

Coefficients of genetic diversity based on nuclear DNA microsatellites—average expected heterozygosity, polymorphism, average number of alleles, and allelic richness—showed little difference among the 4 *C. stegaria* populations studied. There is thus no evidence from any of these measures to suggest local loss of nuclear genetic diversity in any of the populations studied. From mtDNA, nucleotide diversity indices, number of polymorphic sites, and the mean number of pairwise differences were comparable in *C. stegaria* populations from the Green, Licking, and Rolling Fork Rivers but markedly lower in the Clinch River population. These differences between the Clinch River population and the other populations are primarily the result of the absence of L2 from the Clinch River population. Haplotypes of lineages L1 and L2 differ by ~12% to 14%, and these large differences are reflected in the high within-population diversity values seen in the 3 populations from the Ohio River drainage. Haplotype diversity was, however, higher in the Clinch River population compared with the remaining populations, suggesting that the presumed historical event, which may have resulted in the absence of L2, did not affect nucleotide diversity within L1.

### Summary and Future Research

Our results reveal multiple patterns of population differentiation in *C. stegaria*, dependent on the marker used. These patterns are not necessarily conflicting but rather confirm the importance of combining data from multiple markers instead of using partial information from single markers to estimate true population structure. Overall, connectivity in *C. stegaria* most likely follows a stepping-stone model along river courses with little real differentiation among adjacent populations but with the geographically most distant populations (the Ohio and Clinch River populations) showing differentiation. This stepping-stone motif has probably been interrupted by anthropogenic influences in recent decades but is still evident from nuclear microsatellite markers that display high homogeneity across populations. The limited genetic differentiation that was detected among adjacent populations based on lineage L1 could reflect the low effective population size of mtDNA markers, with a concurrent increase in rates of drift compared with nuclear markers. The absence of L2 from the Clinch River population most likely reflects an ancient evolutionary event, with the signature sustained through isolation by distance, resulting from a lack of gene flow between individuals from the 2 extremes of the distribution range of *C. stegaria*.

Future research on genetic connectivity in *C. stegaria* should aim to increase sample sizes for existing sites and to add additional sites to achieve better fine-grained coverage of the remaining extant populations of this species. The first benefit of additional sampling would be to indicate whether L2 is truly absent from *C. stegaria* in the Clinch River drainage and thus confirm or disprove the possible uniqueness of this population. Second, wider sampling would confirm whether the local populations sampled from the Green, Rolling Fork, and Licking Rivers are fully representative of these rivers.

As a preliminary management measure to prevent the swamping of a possibly unique evolutionary lineage of the species in the Tennessee River system, we nevertheless recommend that the Clinch River population be treated as a separate MU. The absence of a mitochondrial lineage from the Clinch would merit the designation and management of populations in the 2 drainages as separate management units by at least some criteria—using an evolutionary paradigm (Waples and Gaggiotti 2006) there is no evidence of reproductive cohesiveness indicative of gene flow. The population is likely smaller than those studied in the Ohio River basin, but it is nevertheless recruiting and viable; therefore, immediate augmentation is not a priority. Long-term exclusion or inclusion of the population from an overall metapopulation management strategy for *C. stegaria,* where any population can act as source or recipient population for artificially induced gene flow, needs to be further investigated and considered in the context of additional biological data. We recommend that fish host specificity studies be conducted to determine whether biological differences exist between specimens in the Clinch and the Ohio River drainage streams. Such future work should also include research on the migration of fish hosts, including anthropogenic restrictions to dispersal and ecological drivers such as habitat preferences, and the possible influence of currents (Hänfling and Weetman 2006]).

## Funding

U.S. Geological Survey Biological Resources Division (00HQRU1573, Work Order 92); Virginia Tech Office of Sponsored Programs Project (432622).

## Acknowledgments

We thank Chris Barnhart and Edwin Miller for supplying the *C. aberti* individuals used in this study, and Monte McGregor and Leroy Koch for their help in collecting individuals of *C. stegaria* from Kentucky. Amplification fragment analysis was done at the Virginia Bioinformatics Institute, Blacksburg, VA. The views expressed in this publication are those of the authors and do not necessarily represent the views of the USFWS.

We thank 3 anonymous referees for recommendations based on the initial version of this manuscript.

## References

Bandelt H-J, Forster P, Rőhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Berg DJ, Cantonwine EG, Hoeh WR, Guttman SI. 1998. Genetic structure of *Quadrula quadrula* (Bivalvia: Unionidae): little variation across large distances. J Shellfish Res. 17:1365–1373.

Birky GW, Maruyama T, Fuerst KP. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. Genetics. 103:513–527.

Burdick RC, White MM. 2007. Phylogeography of the Wabash pigtoe, *Fusconaia flava* (Rafinesque, 1820) (Bivalvia: Unionidae). J Molluscan Stud. 74:367–375.

Buhay JE, Serb JM, Dean CR, Parham Q, Lydeard C. 2002. Conservation genetics of two endangered Unionid bivalve species, Epioblasma florentina Walkeri and E. capsaeformis (Unionidae: Lampsilini). J. Molluscan Stud. 68:385–391.

Crandall KA, Bininda-Edmonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology: an alternative to "evolutionarily significant units.". Trends Ecol Evol. 15:290–295.

Elderkin CL, Christian AD, Vaughn CC, Metcalfe-Smith JL, Berg DJ. 2007. Population genetics of the freshwater mussel, *Amblema plicata* (Say 1817) (Bivalvia: Unionidae): evidence of high dispersal and post-glacial colonization. Conserv Gen. 8:355–372.

Ellsworth DL, Honeycut RL, Silvy NJ, Rittenhouse KD, Smith MH. 1994. Mitochondrial-DNA and nuclear-gene differentiation in North American Prairie grouse (genus *Tympanuchus*). Auk. 111:661–671.

Excoffier L, Laval G, Schneider S. 2005. Arlequin: an integrated software package for population genetics data analysis. Version 3.0. Computational and Molecular Population Genetics Laboratory (CMPG). Berne (Switzerland): Institute of Zoology, University of Berne.

Gaggiotti OE, Vetter RD. 1999. Effect of life history strategy, environmental variability, and overexploitation on the genetic diversity of pelagic fish populations. Can J Fish Aquat Sci. 56:1376–1388.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices. Version 293. Lucerne (Switzerland): Institut de Zoologie et d'Ecologie Animale, Universite de Lausane.

Grobler JP, Jones JW, Johnson NA, Beaty B, Struthers J, Neves RJ, Hallerman EM. 2006. Patterns of genetic differentiation and conservation of the slabside pearlymussel, *Lexingtonia dolabelloides* Lea 1840, in the Tennessee river drainage. J Molluscan Stud. 72:65–75.

Groombridge B, editor. 1994. 1994 IUCN Red List of Threatened Animals. Gland (Switzerland): International Union for the Conservation of Nature.

Haag WR, Rypel AL. 2010 July 1. Growth and longevity in freshwater mussels: evolutionary and conservation implications. Biol Rev. doi: 10.1111/j.1469-185X.2010.00146.x.

Hänfling B, Weetman D. 2006. Concordant genetic estimators of migration reveal anthropogenically-enhanced source-sink population structure in the river sculpin, *Cottus gobio*. Genetics. 173:1487–1501.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Johnson RI. 1980. Zoogeography of North American Unionacea (Mollusca: Bivalvia) North of the Maximum Pleistocene Glaciation. Bull Mus Comp Zool. 149:77–189.

Jones JW, Culver M, David V, Struthers J, Johnson N, Neves RJ, O'Brien SJ, Hallerman EM. 2004. Development and characterization of microsatellite loci in the endangered oystermussel *Epioblasma capsaeformis* (Bivalvia Unionidae). Mol Ecol Notes. 10:1–4.

Jones JW, Neves RJ. 2002. Life history and propagation of the endangered fanshell pearlymussel, Cyprogenia stegaria Rafinesque (Bivalvia: Unionidae). J N Am Benthol Soc. 21:76–88.

Machordom A, Araujo R, Erpenbeck D, Ramos M. 2003. Phylogeography and conservation genetics of endangered European Margaritiferidae (Bivalvia: Unionoidea). Biol J Linn Soc. 78:235–252.

Manel S, Berthier P, Luikart G. 2002. Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conserv Biol. 16:650–659.

Mock KE, Brim-Box JC, Chong JR, Howard JK, Nez DA, Wolf D, Gardner RS. 2010. Genetic structuring in the freshwater mussel *Anodonta* corresponds with major hydrologic basins in the western United States. Mol Ecol. 19:569–591.

Mock KE, Brim-Box C, Miller MP, Downing ME, Hoeh WR. 2004. Genetic diversity and divergence among freshwater mussel (*Anodonta*) populations in the Bonneville Basin of Utah. Mol Ecol. 13:1085–1098.

Moritz C. 1994. Defining "evolutionarily significant units" for conservation. Trends Ecol Evol. 9:373–375.

Moritz C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst Biol. 51:238–254.

Mulvey J, Lydeard C, Pyer DL, Hicks KM, Brim-Box J, Williams JD, Butler RS. 1997. Conservation genetics of North American freshwater mussels *Amblema* and *Megalonaias*. Conserv Biol. 11:868–878.

Orsini L, Huttunen S, Schlötterer C. 2004. A multilocus microsatellite phylogeny of the *Drosophila virilis* group. Heredity. 93:161–165.

Palsbøll PJ, Bérebé M, Allendorf FW. 2006. Identification of management units using population genetic data. Trends Ecol Evol. 22:11–16.

Parmalee PW, Bogan AE. 1998. The freshwater mussels of Tennessee. Knoxville (Tennessee): The University of Tennessee Press.

Posada DA, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. Am J Hum Gen. 67:170–181.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution. 43: 223–225.

Roe KJ, Lydeard C. 1998. Molecular systematics of the freshwater mussel genus *Potamilus* (Bivalvia: Unionidae). Malacologia. 39:195–205.

Rosenberg M. 2003. PASSAGE, version 1: A software package for performing spatial analysis and statistics on biological and other data. Center for Evolutionary Functional Genomics / Biodesign Institue, School of Life Sciences Arizona State University. [cited 2010 Dec 3]. Available from: URL. http://www.passagesoftware.net/.

Ryder OA. 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol Evol. 1:9–10.

Serb JM. 2006. Discovery of genetically distinct sympatric lineages in the freshwater mussel *Cyprogenia aberti* (Bivalvia: Unionidae). J Molluscan Stud. 72:425–434.

Serb JM, Barnhart C. 2008. Congruence and conflict between molecular and reproductive characters when assessing biological diversity in the western fanshell *Cyprogenia aberti* (Bivalvia, Unionidae). Ann Missouri Bot Gard. 95:248–261.

Serb JM, Buhay JE, Lydeard C. 2003. Molecular systematics of the North American freshwater bivalve genus *Quadrula* (Unionidae: Ambleminae) based on mitochondrial *ND1* sequences. Mol Phyl Evol. 28:1–11.

Smouse PE, Long JC. 1992. Matrix correlation analysis in anthropology and genetics. J Phys Anthr. 35:187–213.

Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E. 1995. Male and female mitochondrial DNA lineages in the blue mussel (Mytilus edulis) species group. Mol Biol Evol. 12:735–747.

Turner TF, Trexler JC, Harris JL, Haynes JL. 2000. Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. Genetics. 154:777–785.

U.S. Fish and Wildlife Service. 1990. Endangered and threatened wildlife and plants: designation of the freshwater mussel, the fanshell, as an endangered species. Fed Regist. 55:25591–25595.

U.S. Fish and Wildlife Service. 1991. Fanshell (*Cyprogenia stegaria* (= *C. irrorota*)) recovery plan. Atlanta GA: USFWS 37 pp.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 4:535–538.

Waples RS. 1991. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar Fish Rev. 53: 11–22.

Waples RS. 1995. Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. Am Fish Soc Symp. 17:8–27.

Waples RS, Gaggiotti O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 15:1419–1439.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19:395–420.

Yeh FC, Yang R, Boyle T. 1999. POPGENE, v.1.31. Microsoft Windows based freeware for population genetic analysis. Renewable Resources, Faculty of Agricultural, Environmental and Life Sciences, University of Alberta; [cited 2010 December]. Available from: http://www.ualberta.ca/~fyeh/.

Zouros E, Freeman KR, Ball AO, Pogson GH. 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. Nature. 359:412–414.

Received May 24, 2010; Revised October 12, 2010; Accepted November 3, 2010

Corresponding Editor: Jose Lopez