

SG 50-97

**Genetic variability and relationships of the invasive round and tubenose gobies in the Great Lakes: a comparison with native and introduced Eurasian populations using mitochondrial DNA control region sequences.**

**Alison K. Dillon and Carol A. Stepien\***

*Department of Biology*

*Case Western Reserve University*

*Cleveland, Ohio 44106-7080*

\*Corresponding author and all reprint requests

cas20@po.cwru.edu

216-368-3563

Running Title: Genetic variability and relationships of invasive gobies

**Transactions of the American Fisheries Society**

Submitted August 4, 1998

*Abstract.*--- Population genetic structure and systematic relationships were investigated for two exotic fishes introduced to the Great Lakes in 1990, the round goby *Neogobius melanostomus* and the tubenose goby *Proterorhinus marmoratus*, using DNA sequences from the left domain of the mitochondrial DNA control region. Populations were compared from different regions in the Great Lakes, an introduced population of the round goby from Poland, and a native population of the round goby from the Black Sea. The round goby was characterized by relatively high genetic variability, and 17 haplotypes were identified from 64 individuals. Levels of genetic variability were similar in the invasive and native populations, indicating large founding populations and lack of bottlenecks. The Black Sea was eliminated as a possible founding source for the round goby in both the Great Lakes and the exotic population in Poland. Substitutions in the left domain of the control region revealed significant differences among populations from the Great Lakes and Eurasia, and between pairs of sampling sites in Lakes Erie, Huron, and St. Clair, suggesting non-random mating. No variation was detected in the population of the tubenose goby from the Great Lakes, which has been less successful in terms of spread and population growth. A molecular clock calibration suggested that the genera *Neogobius* and *Proterorhinus* diverged about  $5.2 \pm 1.0$  million years ago, indicating that they may have vicariantly separated from a lineage shared with *Gobius* during the hypothesized isolation of the Paratethys basin from the Mediterranean Tethys Sea.

*Keywords:* Control region, exotic species, founder effect, Genetics, Gobiidae, Great Lakes  
*Neogobius melanostomus, Proterorhinus marmoratus.*

## Introduction

*Background and Current State of Knowledge.*--- The Eurasian round goby *Neogobius melanostomus* and tubenose goby *Proterorhinus marmoratus* (Teleostei: Gobiidae; Fig. 1) were discovered in the St. Clair River (Fig. 2) of the Laurentian Great Lakes in 1990 (Mills et al. 1991; Crossman et al. 1992; Jude et al. 1992). Their larvae were presumably introduced from foreign ballast water discharge (Mills et al. 1991). The repeated successes of nonindigenous species in the Great Lakes have had serious ecological impacts, resulting in many permanent changes (Mills et al. 1993). The introductions of round and tubenose gobies were initially believed to be of little ecological importance (Marsden et al. 1997), unlike previous invasions by the sea lamprey, dreissenid mussels, and the ruffe (Mills et al. 1991, 1993).

Since its introduction, the round goby has spread to all five Great Lakes - faster than any previously introduced fish (Fig. 2; Jude et al. 1992; Marsden et al. 1997). The tubenose goby has not spread as rapidly, but recently colonized northwestern Lake Erie (U.S. Geological Survey 1997). The rapid growth and spread of the round goby throughout the Great Lakes now constitute an ecological concern (e.g., we observed a catch of 513 round goby individuals by the Ohio Division of Wildlife, Fairport Harbor, in a single 15-minute trawl in Lake Erie off Avon Point, Ohio on 10-11-96). The primary concentration of the present study was on the round goby, since it now presents a greater ecological concern. Levels of genetic variability in populations of exotic species are hypothesized to be positively correlated with invasive success (Williamson 1996). Comparing the genetic variability, spread patterns, and founding sources of the two species of gobies with other exotics introduced to the Great Lakes region may help us to understand the common factors regulating their relative successes.

The success of the round goby has been attributed to its wide range of habitats, fecundity, opportunistic diet (including dreissenid mussels), aggressive territorial interactions, and well-developed lateral line sensory system (Marsden et al. 1997). Its reproductive success has been attributed to multiple clutches and male nest guarding (Kovtun 1979; Jude et al. 1995). Some native fishes in the Great Lakes, such as the mottled sculpin *Cottus bairdi*, have declined in areas where the round goby has become abundant, presumably due to competition for food and/or benthic habitat (Crossman et al. 1992; Jude et al. 1995).

*Systematic Relationships and Eurasian Distributions.*--- The genera *Neogobius* and *Proterorhinus* comprise the subfamily Neogobiinae (Simonovic et al. 1996). The neogobiins have fused pelvic fins and elongated dorsal and anal fins (Fig. 1; Miller 1986) and are found in the Ponto-Caspian region (which includes the Marmara, Black, Azov, Caspian, and Aral Seas; Fig. 3). The Ponto-Caspian *Neogobius/Proterorhinus* lineage is hypothesized to have descended from ancestors shared with *Gobius*, that either entered the Black Sea basin earlier during Quaternary interglacial connections with the Mediterranean Sea (Stevanovic 1982) or during the mid-Miocene Epoch (McKay and Miller 1991). *Neogobius* and *Proterorhinus* are distinguished from the Atlantic and Mediterranean basin *Gobius* spp. by two morphological synapomorphies, the absence of a swimbladder and location of the uppermost rays of the pectoral fins within the fin membrane in the former of the two groups (Pinchuk 1991). The monotypic *Proterorhinus* is distinguished from *Neogobius* by the autapomorphy of its long anterior nostrils (Fig. 1B; Miller 1986). The native distribution of the tubenose goby is limited to the Black and Caspian Seas, where it is presently threatened by habitat destruction (Lelek 1989).

The Eurasian distribution of the round goby includes the Sea of Azov, the Sea of Marmara, and the nearshore areas and tributaries of the Black and Caspian Seas (Fig. 3). In

1990, the round goby was discovered in the Gulf of Gdansk of the Baltic Sea (Fig. 3), where it presumably was introduced via ballast water discharge from ships originating in the Black and/or Caspian Seas (Skora and Stolarski 1993). The Gulf of Gdansk had a food base rich in bivalves and lack of large predators in the coastal zone, providing favorable conditions (Skora and Stolarski 1993). In the present study, the population in the Gulf of Gdansk was genetically compared to the invasive population in the Great Lakes and to a native population in the Black Sea.

Dougherty et al. (1996) sequenced 365 base pairs (bp) of the mitochondrial (mt) DNA cytochrome *b* gene from 44 round and two tubenose gobies, finding 63 nucleotide differences between the species ( $p_n = 0.173$ ) and four intraspecific polymorphisms in the round goby ( $p_n = 0.011$ ). Cytochrome *b* evolves relatively slowly in fishes (Kocher et al. 1989; Stepien and Kocher 1997) and the Dougherty et al. (1996) study did not discern enough variation to evaluate differences among sample sites. Dougherty et al. (1996) found five haplotypes (one of which was shared among sites) that each differed by a single nucleotide for the round goby in the Black Sea ( $N=15$ ;  $N=3$  haplotypes) and St. Clair and Detroit Rivers ( $N=29$ ;  $N=3$  haplotypes, one shared with the Black Sea). They concluded that more than one female round goby type was introduced to the Great Lakes basin and that the Black Sea was not the founding source (Dougherty et al. 1996). Dougherty et al. (1996) provided us with their samples, allowing a direct comparison of resolution from the mtDNA control region and cytochrome *b* sequences. An allozyme study found low levels of heterozygosity in a population of the round goby from the Black Sea, Romania ( $N=8$ ,  $P=0.161$ ;  $H=0.021$ ), and five of 31 loci surveyed were polymorphic (Wallis and Beardmore 1984a; 1984b).

*Variation in the mtDNA Control Region.*--- MtDNA is haploid, clonally maternally inherited in most fishes, and does not recombine, facilitating phylogeographic interpretation (Attardi 1985; Avise 1994; Stepien and Kocher 1997). Other population genetic studies of freshwater fishes (Brown et al. 1993; Stepien 1995; Nielsen et al. 1997; Faber and Stepien 1997; Stepien and Faber 1998) found that the left domain of the mtDNA control region, which was sequenced in this study, is the most variable and useful portion.

The control region also is referred to as the displacement loop (D-loop) because during replication, one of the two strands of the helix is displaced by the synthesis of a third strand (Palumbi 1996). The left domain of the control region sometimes contains tandem repeats that are believed to result from replication slippage (Saccone et al. 1991; Madsen et al. 1993) and are found in a variety of fish groups (summarized by Faber and Stepien 1998). The phylogenetic utility of repeated sequence copy number varies and was informative in populations of sturgeon (Brown et al. 1996) and ruffe (Stepien et al. 1998), but lacked signal among walleye (Faber and Stepien 1998). Nucleotide substitutions (point mutations) within the tandem repeats were informative for resolving population genetic relationships for the spotfin shiner (Broughton and Dowling 1997) and walleye (Faber and Stepien 1998).

*Objectives and Hypotheses.*--- The purpose of this investigation was to characterize the genetic variability of exotic round and tubenose goby populations from the Great Lakes using sequences from the left domain of the mtDNA control region. Specific objectives of this investigation were to (1) evaluate the possible number, origin(s) and approximate size(s) of the founding population(s), (2) compare the degree and pattern of genetic variation in native versus introduced areas of their ranges for the round goby, (3) provide a basis for determining whether future sites are the result of spread or new introductions, and (4) compare the relative levels of

intra- versus interspecific genetic variability. Results of this investigation aid understanding of the relative genetic variability and patterns of dispersal of the gobies, in comparison with other exotic species.

## Methods

*Samples Collected.*--- The left domain (~750 bp) was sequenced for 64 round goby individuals from five sites in the Great Lakes (Fig. 2), including; the Shiawassee River, Michigan (site 1, 43° 15' N, 84° 05' W,  $N=5$ ) and the Flint River, Michigan (site 2, 43° 00' N, 83° 45' W,  $N=5$ ) off Lake Huron, Lake St. Clair (site 3, 42° 22' N, 82° 39' W,  $N=10$ ), and Avon Point (site 5, 41° 50' N, 82° 02' W,  $N=10$ ) and the mouth of the Chagrin River, Ohio (site 6, 41° 44' N, 81° 25' W,  $N=10$ ) in Lake Erie. Samples (Fig. 3) of the round goby from Eurasia were sequenced from the Gulf of Gdansk, Poland (site 7, 54° 20' N, 18° 40' E,  $N=20$ ) and the Black Sea, Bulgaria (site 8, 43° 14' N, 27° 58' E,  $N=15$ ). Twelve tubenose goby individuals were analyzed from the St. Clair River (site 4, 42° 45' N, 82° 25' W,  $N=11$ ) and Lake St. Clair (site 3, 42° 26' N, 82° 52' W,  $N=1$ ) in the Great Lakes (Fig. 2). Samples were stored frozen at -80°C or in 95% ethanol at room temperature.

*DNA Extraction, Amplification, and Sequencing.*--- DNA was extracted and purified from the samples following previously described protocols (Stepien 1995). The left domain of the control region, extending from the tRNA-proline gene to the central conserved section (Fig. 4), was amplified with PCR and the oligonucleotide primers L16007 (5'-CCC-AAA-GCT-AAA-ATT-CTA-A-3') for the light chain (Kocher et al. 1989) and H16498 (5'-CCT-GAA-GTA-

GGA-ACC-AGA-TG-3'; Meyer et al. 1990) for the heavy strand. The amplification program for the left domain of the control region was 37 cycles at 94°C for 45 s, 51°C for 40 s, and 72°C for 75 s. Multiple bands were eliminated by excising the ~750 band from a low-melt agarose gel, purification with a GeneClean kit (Bio 101, Inc. #1001-200, Vista, California), and re-amplification. The re-amplification program was 20 cycles at 94°C for 45 s, 54°C for 30 s, and 72°C for 65 s. All PCR programs ended with a final extension step of 72°C for 5 min to ensure that chains were fully polymerized. Samples were sequenced using the Sanger dideoxy-chain termination method (Sanger et al. 1977), run on polyacrylamide gels, and autoradiographed, following Stepien (1995).

*Data Analysis.*--- Proportions of polymorphic nucleotides ( $p_n$ ) were calculated following Nei (1987). Population genetic analysis of variation from mtDNA sequences was conducted in two ways: (1) by analyzing evolutionary relationships among individual haplotypes using genetic distance and cladistic methods and (2) by dividing the data into populations or regions, and testing for geographic heterogeneity and levels of variation within versus among them (Ferris and Berg 1987; Weir 1996).

The first approach compared relationships among individual genotypes using PAUP\* maximum parsimony (Phylogenetic Analysis Using Parsimony; Swofford 1998) and neighbor-joining (NJ) analyses (Saitou and Nei 1987) for 651 bp (bases 90 - 740, Appendix), including all data from the repeated sequences. PAUP\* analyses of relationships among mtDNA haplotypes used the branch-and-bound algorithm (Hendy and Penny 1982), 50% majority rule consensus of the most parsimonious trees (Margush and McMorris 1981), and 1,000 bootstrap replications (Swofford et al. 1996). Kimura (1980) two-parameter and pairwise ( $p$ -) genetic distances (Nei 1987), standard errors, and NJ were calculated with MEGA (Molecular Evolutionary Genetics

Analysis, version 1.01; Kumar et al. 1993). Kimura (1980) two-parameter distances were used to correct for the unequal rate of transitional versus transversional substitutions that occur in animal mtDNA sequences (Kocher and Carlton 1997). P-distances were used to estimate possible divergence times, using a rate of sequence divergence of two percent per million years (reviewed by Avise 1994). This rate is slower than that of mammals (Avise 1994) but average for fishes, whose mtDNA evolves more slowly apparently due to poikilothermy (Bernatchez et al. 1992; Bernatchez and Danzmann 1993; Stepien and Kocher 1997). Bootstrap support from 1,000 permutations was calculated for the nodes of the distance and parsimony trees.

The second approach involved comparing the relative frequencies of haplotypes and the divergences within and among sites, using measures of haplotype diversity ( $h$ ; which is equivalent to heterozygosity; Nei and Tajima 1983) and maximum-likelihood analysis of the average number of nucleotide substitutions per site within (nucleotide diversity-  $\Pi$ ; Nei 1987) and among populations (nucleotide divergence-  $d_{xy}$ ; Nei and Tajima 1983) from the DA2 program in REAP software package (Restriction Enzyme Analysis Package, version 5.0; McElroy et al. 1992; 1997).

Conventional population statistics are inappropriate when too many haplotypes obscure the phylogenetic signal (Bowen and Grant 1997), as occurred in our data set. Therefore, haplotypes were combined in two ways in order to reduce phylogenetic noise, enabling tests for geographic heterogeneity. First, the sequences (including the repeats) were analyzed based on transversional substitutions alone. Transitions accumulate more rapidly than transversions in mtDNA (summarized in Avise 1994) and are often believed to constitute phylogenetic noise (Bowen and Grant 1997). Second, the data set was restricted to the repeated region and the haplotypes were based on transversions (denoted as + in Table 1) and groups that were

distinguished by multiple transitions (denoted as II in Table 1), eliminating singleton transition autapomorphies. Frequency differences in the geographic distributions of both types of combined haplotypes were analyzed with modified chi-square tests, using a Monte Carlo simulation approach with 10,000 randomizations to account for small sample sizes and empty cells in the contingency matrix (Roff and Bentzen 1989) and the MONTE program in REAP 5.0 (McElroy et al. 1997). A Bonferroni correction (Cooper 1968) that divided the *P*-value by the number of pairwise comparisons was used for the multiple post-hoc tests (Fry et al. 1993).

Hierarchical AMOVA (Analysis of MOlecular VAriance, version 1.53; Excoffier 1995) analyses, based on squared Euclidean distances among pairs of haplotypes and their relative frequencies, tested the partitioning of variation between regions (North America and Europe), and among sampling sites. Significance of the  $\Phi_{st}$  ( $F_{ST}$  analog) values was determined from comparison with 1,000 random permutations at the 0.05 confidence level (Excoffier et al. 1992).

Tajima's (1989) test of selective neutrality based on the infinite-site model was performed on the round goby samples from Lakes Erie, Huron, and St. Clair, and Poland and the Black Sea using the Arlequin software package (version 1.1; Schneider et al. 1997). Tajima's (1989) test evaluated whether the observed number of polymorphic sites and the average number of pairwise nucleotide differences deviated significantly from theoretically predicted expected values (Tajima 1989). A separate analysis was run with all haplotypes considered as a single population.

## Results

*Structure and Substitutions in the Left Domain of the mtDNA Control Region.*--- The consensus sequences for *N. melanostomus* and *P. marmoratus* are reported in the Appendix. Inter- and intraspecific substitutions were mapped on a structural diagram of the left domain of

the mtDNA control region (Fig. 4). No intraspecific variation was found in the tubenose goby ( $N=12$ ). The single tubenose goby haplotype (Appendix) and the five most common round goby haplotypes (aa, cc, dd, ee, and gg; Table 2) were reported in GenBank (accession numbers - through - ).

Two putative Termination Associated Sequences (TAS, Doda et al. 1981; Foran et al. 1988) were identified in each of the species (Appendix). The region immediately flanking the 5' tRNA-proline end (bases 1-89; Appendix) differed in sequence and length between the species, and was omitted in all analyses due to alignment uncertainty. The left domain of the round goby was approximately 400 bp longer than that of the tubenose goby, due to four non-identical tandemly repeated sequences (Table 1, Fig. 4). The repeated sequences were A+T rich (>75%) and each contained three TAS related motifs (Table 1). A fifth shortened version (26 bp) of the repeat sequence immediately followed the repeat section, and also was found in the tubenose goby (Appendix, Fig. 4).

*Interspecific Differences.*--- The round and tubenose gobies differed by 26 transitional (S) and 23 transversional (V) substitutions within 306 bp of aligned sequence data (excluding the four repeats; S:V ratio = 1.13; Appendix). Nine insertion/deletions encompassing a total of 50 nucleotide positions occurred between them, including four in the A+T rich region and five in the non-repetitive sequences (Appendix). Mean sequence divergences between *N. melanostomus* and *P. marmoratus* were Kimura (1980) two parameter  $d = 0.112 \pm 0.024$  and pairwise  $p = 0.104 \pm 0.020$ .

*Intraspecific Variation in the Round Goby.*--- Intraspecific variations in the round goby occurred in the repeated sequence array (Table 1) and the non-repetitive sequences (Table 2). Sequence similarity among the repeat variants (A through K) ranged from 87% (between pairs

A-K and B-H) to 99% (between pairs A-C, A-D, F-G, H-J, and H-I), equivalent to  $d = 0.125$  and  $0.010$ , respectively. Thirteen transitions (S), four transversions (V), and a two base insertion occurred within the four tandem repeats (S:V ratio = 3.25; Table 1).

The first three repeats of the four repeat assembly were 106 bp each (Fig. 4), and contained seven variants (designated A - G in Table 1). The variants differed by a total of ten transitional substitutions and  $d = 0.010$  (between A-C, A-D, and F-G) to  $0.082$  (between B-F). The fourth repeat was 108 bp (Fig. 4) and had four variants (H - K). Variants of the fourth repeat differed by four transitions, one transversion, and  $d = 0.009$  (between H-I and H-J) to  $0.039$  (between I-K). The assembly of the four repeated units varied among individuals (Table 2).

MtDNA control region sequence data, including all repeat information, revealed 17 haplotypes of *N. melanostomus* and 43 polymorphic nucleotide sites ( $p_n = 0.066$ ; S:V ratio = 13.0 Table 2, Fig. 4). Nucleotide polymorphism within the four repeats ( $p_n = 0.075$ ) was higher than in the non-repetitive left domain ( $p_n = 0.049$ ). Intrapopulational polymorphisms were detected in five of the seven populations surveyed (the Shiawassee and Flint Rivers of Lake Huron were not polymorphic; Table 3). Nucleotide polymorphism ( $p_n$ ), haplotype diversity ( $h$ ), and nucleotide diversity ( $\Pi$ ) values within the sampling sites of *N. melanostomus* are given in Table 3. These values were largest in the native population from the Black Sea ( $p_n = 0.046$ ;  $h = 0.962 \pm 0.041$ ) and the Lake St. Clair population ( $\Pi = 0.012 \pm 0.007$ ; Table 3). Haplotype diversity ( $h$ ) averaged  $0.458 \pm 0.019$  overall. Nucleotide divergences ( $d_{xy}$ ) among pairs of sampling sites ranged from  $0.0002$  (between Poland and the Black Sea) to  $0.0185$  (between Poland and Lake Huron). Within the three Great Lakes,  $d_{xy}$  was smallest between Lakes Huron and St. Clair ( $0.0009$ ), intermediate between Lakes Erie and St. Clair ( $0.0015$ ), and considerably larger between Lakes Erie and Huron ( $0.0060$ ).

the mtDNA control region (Fig. 4). No intraspecific variation was found in the tubenose goby ( $N=12$ ). The single tubenose goby haplotype (Appendix) and the five most common round goby haplotypes (aa, cc, dd, ee, and gg; Table 2) were reported in GenBank (accession numbers - through - ).

Two putative Termination Associated Sequences (TAS, Doda et al. 1981; Foran et al. 1988) were identified in each of the species (Appendix). The region immediately flanking the 5' tRNA-proline end (bases 1-89; Appendix) differed in sequence and length between the species, and was omitted in all analyses due to alignment uncertainty. The left domain of the round goby was approximately 400 bp longer than that of the tubenose goby, due to four non-identical tandemly repeated sequences (Table 1, Fig. 4). The repeated sequences were A+T rich (>75%) and each contained three TAS related motifs (Table 1). A fifth shortened version (26 bp) of the repeat sequence immediately followed the repeat section, and also was found in the tubenose goby (Appendix, Fig. 4).

*Interspecific Differences.*--- The round and tubenose gobies differed by 26 transitional (S) and 23 transversional (V) substitutions within 306 bp of aligned sequence data (excluding the four repeats; S:V ratio = 1.13; Appendix). Nine insertion/deletions encompassing a total of 50 nucleotide positions occurred between them, including four in the A+T rich region and five in the non-repetitive sequences (Appendix). Mean sequence divergences between *N. melanostomus* and *P. marmoratus* were Kimura (1980) two parameter  $d = 0.112 \pm 0.024$  and pairwise  $p = 0.104 \pm 0.020$ .

*Intraspecific Variation in the Round Goby.*--- Intraspecific variations in the round goby occurred in the repeated sequence array (Table 1) and the non-repetitive sequences (Table 2). Sequence similarity among the repeat variants (A through K) ranged from 87% (between pairs

A-K and B-H) to 99% (between pairs A-C, A-D, F-G, H-J, and H-I), equivalent to  $d = 0.125$  and  $0.010$ , respectively. Thirteen transitions (S), four transversions (V), and a two base insertion occurred within the four tandem repeats (S:V ratio = 3.25; Table 1).

The first three repeats of the four repeat assembly were 106 bp each (Fig. 4), and contained seven variants (designated A - G in Table 1). The variants differed by a total of ten transitional substitutions and  $d = 0.010$  (between A-C, A-D, and F-G) to  $0.082$  (between B-F). The fourth repeat was 108 bp (Fig. 4) and had four variants (H - K). Variants of the fourth repeat differed by four transitions, one transversion, and  $d = 0.009$  (between H-I and H-J) to  $0.039$  (between I-K). The assembly of the four repeated units varied among individuals (Table 2).

MtDNA control region sequence data, including all repeat information, revealed 17 haplotypes of *N. melanostomus* and 43 polymorphic nucleotide sites ( $p_n = 0.066$ ; S:V ratio = 13.0 Table 2, Fig. 4). Nucleotide polymorphism within the four repeats ( $p_n = 0.075$ ) was higher than in the non-repetitive left domain ( $p_n = 0.049$ ). Intrapopulation polymorphisms were detected in five of the seven populations surveyed (the Shiawassee and Flint Rivers of Lake Huron were not polymorphic; Table 3). Nucleotide polymorphism ( $p_n$ ), haplotype diversity ( $h$ ), and nucleotide diversity ( $\Pi$ ) values within the sampling sites of *N. melanostomus* are given in Table 3. These values were largest in the native population from the Black Sea ( $p_n = 0.046$ ;  $h = 0.962 \pm 0.041$ ) and the Lake St. Clair population ( $\Pi = 0.012 \pm 0.007$ ; Table 3). Haplotype diversity ( $h$ ) averaged  $0.458 \pm 0.019$  overall. Nucleotide divergences ( $d_{xy}$ ) among pairs of sampling sites ranged from  $0.0002$  (between Poland and the Black Sea) to  $0.0185$  (between Poland and Lake Huron). Within the three Great Lakes,  $d_{xy}$  was smallest between Lakes Huron and St. Clair ( $0.0009$ ), intermediate between Lakes Erie and St. Clair ( $0.0015$ ), and considerably larger between Lakes Erie and Huron ( $0.0060$ ).

Sequences other than the four tandem repeats revealed eight transitions (S), two transversions (V), and a single insertion (S:V ratio = 4.0). Most of this variation (ten of eleven polymorphic sites) occurred in the Eurasian populations (Table 2). A single transition occurred in the fifth shortened repeat (SR), found in three individuals from both Eurasian sites (Appendix, Fig. 4). Two of the transitions (bases 575 and 588) occurred in single individuals sampled from the Black Sea (Table 2). Nucleotide position 671 was fixed in both of the Eurasian populations (G), and a transversion (53% T- 47% G) occurred in some North American samples.

Restricted haplotypes were calculated in two different ways, enabling statistical tests for geographical heterogeneity among sampling sites (see Methods). The reduced haplotype data sets are given in Table 4. Variance components and associated  $\Phi$  statistics showed significant hierarchical geographic partitioning of genetic variation using both types of restricted haplotype data in AMOVA tests (see Tables 4 and 5; Excoffier 1995). Most of the variation occurred within the sampling sites (57.50%, Table 5A). Variation between the North American and Eurasian regions (20.85%) was similar to that among the sampling sites within the regions (21.64%; Table 5A). All were statistically significant (Table 5).

Modified chi-square tests (Roff and Bentzen 1989; McElroy et al. 1992) showed significant differences in the distributions of the restricted haplotypes (combined entire data set; Table 4A) between North America and Eurasia and among the five sampling sites overall (Table 6A). Tests based on the restricted repeat data (Table 4B) revealed significant differences between North America and Eurasia, among the five sampling sites overall, and among the three Great Lakes (Table 6B), indicating greater geographic heterogeneity. Pairwise comparisons between Lakes Erie, Huron, and St. Clair showed no geographic differences from the restricted

haplotypes based on the entire sequences, and revealed significant structuring in the repeat data (Table 7).

Tajima's (1989) test supported selective neutrality ( $D = -0.23$ ;  $P = 0.43$ ) when all round goby samples were analyzed as a single population. When the data were analyzed as separate sampling sites, selective neutrality was supported in four of the five sampling sites (Lake Erie,  $D = -0.16$ ,  $P = 0.46$ ; Lake Huron,  $D = 2.42$ ,  $P = 0.99$ ; Lake St. Clair,  $D = 1.45$ ,  $P = 0.91$ ; and Poland,  $D = 0.27$ ,  $P = 0.61$ ). Results were statistically significant, rejecting the selective neutrality hypothesis for the population from the Black Sea ( $D = -1.80$ ,  $P = 0.02$ ).

*Genetic Divergences.*--- Kimura (1980) two-parameter distances and p-distances among *N. melanostomus* haplotypes were similar, with overlapping standard errors. NJ analysis produced the tree shown in Figure 5. Haplotype ll from the Black Sea was basal to all other *N. melanostomus*, due to its unique repeat sequences (Tables 1 and 2). The remaining haplotypes formed two subgroups, one containing the majority of the North American haplotypes (83% of the haplotypes and 97% of the individuals; designated as I), and the other comprising mainly Eurasian types (90% exclusively Eurasian haplotypes and 92% of the Eurasian individuals; designated as II; Fig. 5). Group I had three major types, including two from Lakes Erie and St. Clair (type 1, haplotypes aa and bb; type 3, haplotype ee) and another found in Lakes Erie, Huron, and St. Clair (type 2, haplotypes cc and dd). One Black Sea haplotype (qq) clustered with the North American type 3 haplotypes. Type four in Lake St. Clair also was identified in the Black Sea, and was the sister group to a cluster of exclusively Eurasian haplotypes.

Distances among the haplotypes ranged from  $d = 0.002$  (between haplotype pairs aa-bb, ff-gg, gg-jj, gg-oo, ii-oo, nn-oo, and oo-pp) to 0.042 (between haplotype pair ll-dd; Fig. 5), averaging  $0.014 \pm 0.005$ . Haplotypes in group I (Fig. 5) were separated by an average genetic d

=  $0.009 \pm 0.004$ , and group II by  $d = 0.006 \pm 0.003$ . The four types identified in the North American population (Fig. 5) diverged by a mean  $d = 0.011 \pm 0.004$ . Molecular clock calibrations estimated that these divergences occurred during the mid-Pleistocene, ranging from  $300 \pm 150$  (within group II) to  $550 \pm 200$  thousand years ago (among the four types in North America).

PAUP\* analysis (Swofford 1998) found 12 most parsimonious trees based on the entire data set, having 497 steps and consistency indices (excluding uninformative characters) of 0.86. The 50% majority rule consensus tree is shown in Figure 6. Several nodes were unresolved due to the prevalence of single point mutations separating haplotypes. The North American types were contained in a single clade (denoted as A), along with some individuals from the Black Sea. Most haplotypes from the Black Sea were contained in clade B.

## Discussion

*Structure and Substitutions in the mtDNA Control Region.* --- Repeated mtDNA control region sequences were useful for resolving phylogeographic relationships among populations of the round goby. The repeat units were long, contained TAS related elements, and formed energetically stable secondary structures similar to those found in sturgeon (Buroker et al. 1990; Miracle and Campton 1995; Brown et al. 1996) and flatfishes (Lee et al. 1995). In contrast, repeated sequences in percids (Faber and Stepien 1997; 1998; Stepien et al. 1998) lacked TAS motifs. The formation of repeats in the round goby and other groups appears consistent with the hypothesis of slipped strand mispairing during replication and repair, resulting in the addition or loss of repeats (Levinson and Gutman 1987). Slipped strand mispairing is believed to produce

repeated sequences in the mitochondrial and nuclear genomes, regulated by concerted evolutionary processes (Levinson and Gutman 1987; Broughton and Dowling 1997).

In contrast with repeated motifs in flatfishes (Lee et al. 1995) and percids (Faber and Stepien 1998; Stepien et al. 1998), many nucleotide polymorphisms occurred in the four repeats of *N. melanostomus*. While some nucleotide substitutions were homogenized across the repeat array; others appeared to be incompletely homogenized, indicating an intermediate stage in the process of concerted evolution (Fumagalli et al. 1996). Walleye had a final repeat variant with a high number of substitutions among individuals, that also appeared to lack homogenization (Faber and Stepien 1997; 1998). High variation in the repeated sequences of the round goby may be explained by different balances between point mutations and the homogenization of insertion/deletion events (Fumagalli et al. 1996) in various groups of fishes.

Tests for selective neutrality supported that polymorphisms in the mtDNA control region of the round goby are neutral mutations. A single deviation in the Black Sea population may be due to sampling error and small sample size, lack of random mating, and/or selection (Tajima 1989).

*Interspecific Variation Between Round and Tubenose Gobies.*--- The round and tubenose gobies were separated by an average sequence p-distance of  $0.104 \pm 0.020$ , corresponding to about  $5.2 \pm 1.0$  million years (using a molecular clock calibration; see Methods). *Neogobius* and *Proterorhinus* may have diverged during the late Miocene to early Pliocene Epochs, refuting the hypothesis that the neogobiin Ponto-Caspian lineage diverged from *Gobius* during Quaternary interglacial connections between the basins of the Black and Mediterranean Seas (Stevanovic 1982). McKay and Miller (1991) suggested that the groups diverged during the mid-Miocene, based on isozyme analysis, which is more similar to the genetic divergence in our study. The

*Gobius* and neogobiin lineages may have vicariantly separated via isolation of the Paratethys from the Mediterranean Tethys Sea basins, about 10-12 million years ago (Steininger and Rogl 1984). Our study suggests that the neogobiins diverged from the *Gobius* lineage during the mid-Miocene, and then speciated into *Neogobius* and *Proterorhinus* during the late Miocene/early Pliocene Epochs.

*Intraspecific Divergences.*--- Levels of haplotypic and nucleotide diversity for the round goby were similar in samples from Lake St. Clair (the North American founding population) and a native population from the Black Sea (Table 3), indicating that the population introduced to North America was relatively large and did not undergo a genetic bottleneck. The Lake St. Clair population was less variable in the non-repetitive sequence, and had higher diversity within repeated sequences, in comparison to the Black Sea population. In contrast, levels of genetic diversity were somewhat lower in the introduced population in the Gulf of Gdansk, Poland, suggesting a possible founder effect (Table 3).

A single haplotype was shared between the Lake St. Clair and Black Sea populations (17% of the individuals), indicating that the North American population did not originate from the Black Sea. This corroborates the results of the cytochrome *b* study by Dougherty et al. (1996). A single haplotype (gg) was shared between the Black Sea and the introduced population in Poland, indicating that the Black Sea also was not its founding source. The exotic populations in Poland and the Great Lakes were probably founded by independent sources, as our study found no shared haplotypes (Table 2).

The presence of four widely diverged types (Fig. 5) in the North American population indicates that a large number of individuals were introduced from one or more unknown source populations. Divergence time estimates for the four types suggest their divergence during the

mid-Pleistocene, presumably due to vicariant isolation in glacial refugia (Hewitt 1996). Other freshwater fishes in Europe exhibit similar levels of genetic divergences related to glacial refugia (Bernatchez and Osinov 1995; Hanfling and Brandl 1997; Stepien et al. 1998). Differences among sites in the Great Lakes appear consistent with multiple founding sources, low gene flow, and/or lineage extinction.

All haplotypes from Lakes Erie and Huron were present in the Lake St. Clair site (the point of introduction), except haplotype bb ( $N=3$ ; Lake Erie). The presence of a unique haplotype in Lake Erie may be due to an independent introduction or to sampling error. Five of the six Great Lakes haplotypes were found in Lake St. Clair, suggesting that the other sites are the result of intra- and inter-lake spread. Sampling locations in Lakes Erie and Huron (presumably areas of spread) had fewer haplotypes and lower haplotypic diversity (Tables 2 and 3), potentially due to founder effects. The two tributaries of Lake Huron each contained a single haplotype, suggesting that they were founded by different genetic types and/or experienced differential lineage extinction. It has been hypothesized that the Lake Huron tributaries may have originated from transport of bait fish from Lake St. Clair (D. Jude, University of Michigan, pers. commun. May 1998). Results are consistent with the hypothesis of separate bait fish introductions, as both genetic types were found in Lake St. Clair.

In contrast to the cytochrome *b* study by Dougherty et al. (1996), mtDNA control region sequences showed greater variability and significant differences among sampling sites of the round goby. Unlike genetic studies of other recently introduced exotics in the Great Lakes (e.g., zebra mussels, Marsden et al. 1996; ruffe, Stepien et al. 1998), the round goby showed significant geographical structuring among Lakes Erie, St. Clair, and Huron, indicating barriers to gene flow and/or non-random mating (Table 7). Lack of a pelagic larval stage (Marsden et al.

1997) may limit gene flow among sites. The male round goby is aggressive, actively recruits mates, and defends the nest (Jude 1992). Reports indicate that males reproduce only once at a large size (about age 3), do not feed during nest-guarding, and then may die (Marsden et al. 1997). This indicates a highly selective strategy and a substantial fitness investment in a single reproduction.

All tubenose goby individuals surveyed from the St. Clair region were represented by a single haplotype (Appendix), suggesting that either the source population had low genetic variability or that the founding population was small. The maternally inherited mtDNA genome has a small effective population size ( $\frac{1}{4}$  the size of the bisexually inherited diploid nuclear genome), increasing its sensitivity to bottlenecks and founding events (Birky et al. 1989; Avise 1994). Investigations of native Eurasian tubenose goby populations are necessary to address this question.

The round goby has undergone a more rapid expansion in the Great Lakes and has a higher level of genetic variability than the tubenose goby. Differences in relative success between the two gobies also may be related to the larger size and more aggressive nature of the round goby (Jude et al. 1995; Marsden et al. 1997). Expansion of the round goby population may be augmented by the availability of its zebra mussel prey, for which there are few competitors (Jude 1997). The smaller tubenose goby primarily consumes smaller prey, including aquatic insects and benthic zooplankton, for which there are more native competitors (Jude et al. 1995). The present study provides a baseline data set that will allow populations of round and tubenose gobies to be genetically monitored, in order to elucidate whether there are additional introductions and evaluate their future patterns of spread.

## **Acknowledgments**

This study was supported by grants to C. Stepien from the Ohio Sea Grant Program development fund #RF-729753 (1997-98), and the Lake Erie Protection Fund #SG 50-97 (1997-98). Specimens were provided by the Michigan Department of Natural Resources (Bob Haas and Mike Thomas), the Ohio Division of Wildlife (Carey Knight and Kevin Kayle), David Jude (University of Michigan), Jeffrey Ram (Wayne State University), Svetlana Rudnicka, and Mariusz Sapoto (Institute of Oceanography, Poland). P. J. Miller gave permission to use illustrations in Figure 3, adapted from Miller (1986). The Illinois Natural History Survey and Illinois-Indiana Sea Grant Program (C. Ward) and Ellen Marsden (University of Vermont) gave permission to adapt Figure 3 from Marsden et al. (1997).

## Literature Cited

- Attardi, G. 1985. Animal mitochondrial DNA: an extreme example of genetic economy. *International Review of Cytology* 93:93-145.
- Awise, J. C. 1994. *Molecular markers, natural history and evolution*. Chapman and Hall, New York.
- Bernatchez, L., and Danzmann, R. G. 1993. Intraspecific DNA sequence variation of the mitochondrial DNA of brook char (*Salvelinus fontinalis* Mitchell). *Molecular Biology and Evolution* 10:1002-1014.
- Bernatchez, L., Guyomard, R., and Bonhomme, F. 1992. DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology* 1:161-173.
- Bernatchez, L., and Osinov, A. 1995. Genetic diversity of trout (genus *Salmo*) from its most eastern native range based on mitochondrial DNA and nuclear gene variation. *Molecular Ecology* 4:285-297.
- Birky, C. W., Fuerst, P., and Maruyama, T. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectation, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121:613-627.
- Bowen, B. W., and Grant, W. S. 1997. Phylogeography of the sardines (*Sardinops* spp.): assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* 51:1601-1610.

- Broughton, R. E., and Dowling, T. E. 1997. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella spiloptera*. *Molecular Biology and Evolution* 14:1187-1196.
- Brown, J. R., Beckenbach, A. T., and Smith, M. J. 1993. Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). *Molecular Biology and Evolution* 10:326-341.
- Brown, J. R., Beckenbach, K., Beckenbach, A. T., and Smith, M. J. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). *Genetics* 142:525-535.
- Buroker, N. E., Brown, J. R., Gilbert, T. A., O'Hara, P. J., Beckenbach, A. T., Thomas, W. K., and Smith, M. J. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* 124:157-163.
- Cooper, D. W. 1968. The significance level in multiple tests made simultaneously. *Heredity* 23:614-617.
- Crossman, E. J., Holm, E., Cholmondeley, R., and Tuininga, K. 1992. First record for Canada of the rudd, *Scardinius erythrophthalmus*, and notes on the introduced round goby, *Neogobius melanostomus*. *The Canadian Field-Naturalist* 106:206-209.
- Doda, J. N., Wright, C. T., and Clayton, D. A. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proceedings of the National Academy of Sciences of the USA* 78:6116-6120.
- Dougherty, J. D., Moore, W. S., and Ram, J. L. 1996. Mitochondrial DNA analysis of round goby (*Neogobius melanostomus*) and tubenose goby (*Proterorhinus marmoratus*) in the Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences* 53:474-480.

- Excoffier, L. 1995. AMOVA, Analysis of Molecular Variance, version 1.55. University of Geneva, Switzerland.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA data. *Genetics* 131:479-491.
- Faber, J. E., and Stepien, C. A. 1997. The utility of mitochondrial DNA control region sequences for analyzing phylogenetic relationships among populations, species, and genera of the Percidae. Pages 129-143 in T. D. Kocher and C. A. Stepien, editors. *Molecular systematics of fishes*. Academic Press, San Diego, California.
- Faber, J. E., and Stepien, C. A. 1998. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedion*. *Molecular Phylogenetics and Evolution*, in press.
- Ferris, S. D., and Berg, W. J. 1987. The utility of mitochondrial DNA in fish genetics and fishery management. Pages 277-299 in N. Ryman and F. Utter, editors. *Population genetics and fishery management*. University of Washington Press, Seattle, Washington.
- Foran, D. R., Hixson, J. E. and Brown, W. M. 1988. Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Research* 16:5841-5861.
- Fry, J. C., Iles, T. C., Randerson, P. F., Bridge, P. D., Dunstan, F. D. J., Bowker, D. W., and Weingert, R. G. 1993. *Biological data analysis: a practical approach*. Oxford University Press, New York.

- Fumagalli, L., Taberlet, P., Favre, L., and Hausser, J. 1996. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Molecular Biology and Evolution* 13:31-46.
- Hanfing, B. and Brandl, R. 1997. Genetic differentiation of the bullhead *Cottus gobio* L. across watersheds in Central Europe: evidence for two taxa. *Heredity* 80:110-117.
- Hendy M. D., and Penny, D. 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Bioscience* 59:277-290.
- Hewitt, G. M. 1996. Some genetic consequences of Ice Ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society, London* 58:247-27.
- Jude, D. J. 1997. Round gobies: Cyberfish of the third millennium. *Great Lakes Research Review* 3:27-34.
- Jude, D. J., Janssen, J., and Crawford, G. 1995. Ecology, distribution, and impact of the newly introduced round and tubenose gobies in the biota of the St. Clair and Detroit Rivers. Pages 447-460 in M. Munawar, T. Edsall, and J. Leach, editors. *The Lake Huron ecosystem: ecology, fisheries and management*. SPB Academic Publishing, Amsterdam.
- Jude, D. J., Reider, R. H., and Smith, R. 1992. Establishment of Gobiidae in the Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences* 49:416-421.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.
- Kocher, T. D., and Carlton, K. L. 1997. Base substitution in fish mitochondrial DNA: patterns and rates. Pages 13-24 in T. D. Kocher and C. A. Stepien, editors. *Molecular systematics of fishes*. Academic Press, San Diego, California.

- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X., and Wilson, A. C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA* 86:6196-6200.
- Kovtun, I. F. 1979. Significance of the sex ratio in the spawning population of the round goby in relation to year-class strength in the Sea of Azov. *Journal of Ichthyology* 19:161-163.
- Kumar, S., Tamura, K., and Nei, M. 1993. *Molecular Evolutionary Genetics Analysis*, version 1.02. Institute of Molecular Evolutionary Genetics, University of Pennsylvania, University Park, Pennsylvania.
- Lee, W., Conroy, J., Howell, W. H., and Kocher, T. D. 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution* 41:54-66.
- Lelek, A. 1989. Threatened fishes of Europe, volume 9 *in* Council of Europe: European Committee for the conservation of nature and natural resources, editors. The freshwater fishes of Europe. AULA Verlag, Wiesbaden.
- Levinson, G., and Gutman, G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4:203-221.
- Madsen, C. S., Ghivizzani, S. C. and Hauswirth, W.W. 1993. *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. *Proceedings of the National Academy of Sciences of the USA* 90:7671-7675.
- Margush, T. and McMorris F. R. 1981. Consensus n-trees. *Bulletin of Mathematical Biology* 42:239-244.

- Marsden, J. E., Charlebois, P., Wolfe, K., Jude, D., and Rudnicka, S. 1997. The round goby (*Neogobius melanostomus*): a review of European and North American literature with notes from the Round Goby Conference, Chicago. Illinois Natural History Survey. Special Publication Number 20, Champaign, Illinois.
- Marsden, J. E., Spidle, A., and May, B. 1996. Review of genetic studies of *Dreissena* spp. *American Zoologist* 36:259-270.
- McElroy, D. P., Moran, P., Bermingham, P. E., and Kornfield, I. 1992. REAP: An integrated environment for the manipulation and phylogenetic analysis of restriction data. *Journal of Heredity* 83:157-158.
- McElroy, D. P., Moran, P., Bermingham, P. E., and Kornfield, I. 1997. REAP, Restriction Enzyme Analysis Package, version 5. Western Kentucky University, Bowling Green, Kentucky.
- McKay, S. I., and Miller, P. J. 1991. Isozyme criteria in the testing of phyletic relationships between species of *Gobius* and related eastern Atlantic-Mediterranean genera (Teleostei: Gobiidae) *Journal of Fish Biology* 39 (Suppl. A):291-299.
- Meyer, A., Kocher, T. D., Basatbwaki, P., and Wilson, A. C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347:550-553.
- Mills, E. L., Leach, J. H., Carlton, J. T., and Secor, C. L. 1991. Exotic species in the Great Lakes: a history of biotic crises and anthropogenic introductions. Great Lakes Fishery Commission, Ann Arbor, Michigan.

- Mills, E. L., Leach, J. H., Carlton, J. T. and Secor, C. L. 1993. Exotic species in the Great Lakes: a history of biotic crises and anthropogenic introductions. *Journal of Great Lakes Research* 19:1-54.
- Miller, P. J. 1986. Gobiidae. Pages 1019-1095 in P. J. P. Whitehead, M. L. Bauchot, J. C. Hureau, J. Nielsen, and E. Tortonese, editors. *Fishes of the northeast Atlantic and Mediterranean*. UNESCO, Paris.
- Miracle, A. L., and Campton, D. E. 1995. Tandem repeats sequence variation and length heteroplasmy in the mitochondrial DNA d-loop of the threatened Gulf of Mexico sturgeon, *Acipenser oxyrinchus desotoi*. *Journal of Heredity* 86:22-27.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei, M., and Tajima, F. 1983. Maximum likelihood estimation of the number of nucleotide substitution from restriction-site data. *Genetics* 105:207-217.
- Nielsen, J. L., Fountain, M. C., and Wright, J. M. 1997. Biogeographic analysis of Pacific trout (*Oncorhynchus mykiss*) in California and Mexico based on mitochondrial DNA and nuclear microsatellites. Pages 53-74 in T. D. Kocher and C. A. Stepien, editors. *Molecular systematics of fishes*. Academic Press, San Diego, California.
- Palumbi, S. R. 1996. Nucleic acids II: the polymerase chain reaction. Pages, 205-221 in D. M. Hillis, C. Moritz, and B. K. Mable, editors. *Molecular systematics*, 2nd edition. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Pinchuk, V. I. 1991. K voprosy grupirovokakh vidov v predelakh roda *Neogobius* (Perciformes) (About speciation within the genus *Neogobius* (Perciformes)). *Voprosy Ikhtiologii* 31:380-393.

- Roff, D. A., and Bentzen, P. 1989. The statistical analysis of mitochondrial DNA polymorphisms:  $X^2$  and the problem of small samples. *Molecular Biology and Evolution* 6:539-545.
- Saccone, C., Graziano, P., and Sbisà, E. 1991. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. *Journal of Molecular Evolution* 33:83-101.
- Saitou, N., and Nei, M. 1987. The neighbor joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the USA* 74:5463-5467.
- Schneider, S., Kueffer, J., Roessli, D., and Excoffier, L. 1997. Arlequin version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Simonovic, P. D., Nikolic, V. P., and Skora, K. E. 1996. Vertebral number in Ponto-Caspian gobies: phylogenetic relevance. *Journal of Fish Biology* 49:1027-1029.
- Skora, K. E. and Stolarski, J. 1993. New fish species in the Gulf of Gdansk, *Neogobius* spp. [cf. *Neogobius melanostomus* (Pallas 1811)]. *Bulletin of the Sea Fisheries Institute* 1:83.
- Steininger, F. F., and Rogl, R. 1984. Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. Pages 659-686 in J. E. Dixon and A. H. F. Robertson, editors. *The geological evolution of the eastern Mediterranean*. Blackwell, Oxford.

- Stepien, C. A. 1995. Population genetic divergence and geographic patterns from DNA sequences: examples from marine and freshwater fishes. Pages 263-287 in J. L. Nielsen, editor. Evolution and the aquatic ecosystem: defining unique units in population conservation. American Fisheries Society, Symposium 17, Bethesda, Maryland.
- Stepien, C. A., Dillon, A. K., and Chandler, M. D. 1998. Genetic identity, phylogeography, and systematics of ruffe *Gymnocephalus* in the North American Great Lakes. *Journal of Great Lakes Research* 24:361-378.
- Stepien, C. A., and Faber, J. E. 1998. Population genetic structure, phylogeography, and spawning philopatry in walleye (*Stizostedion vitreum*) from mtDNA control region sequences. *Molecular Ecology*, in press.
- Stepien, C. A., and Kocher, T. D. 1997. Molecules and morphology in studies of fish evolution. Pages 1-11 in T. D. Kocher and C. A. Stepien, editors. *Molecular systematics of fishes*. Academic Press, San Diego, California.
- Stevanovic, P. M. 1982. *Istorijska Geologija-Kenozoik (Geological History-Cenozoic)*. Belgrade: University of Belgrade.
- Swofford, D. L. 1998. PAUP\* (Phylogenetic Analysis Using Parsimony) 4.0, Sunderland, Massachusetts: Sinauer Assoc., Inc.
- Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. 1996. Phylogenetic inference. Pages 407-514 in D. M. Hillis, C. Moritz, and B. K. Mable, editors. *Molecular systematics*, 2nd edition. Sinauer Assoc., Inc., Sunderland, Massachusetts.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:858-595.

- U. S. Geological Survey (Biological Resources Division). 1997. Online (range maps).  
[http://nas.er.usgs.gov/fishes/images/goby\\_map.gif](http://nas.er.usgs.gov/fishes/images/goby_map.gif) (round goby)  
[http://nas.er.usgs.gov/fishes/images/pr\\_marmo.gif](http://nas.er.usgs.gov/fishes/images/pr_marmo.gif) (tubenose goby)
- Wallis, G. P., and Beardmore, J. A. 1984a. Genetic variation and environmental heterogeneity in some closely related goby species. *Genetica (The Hague)* 62:223-238.
- Wallis, G. P., and Beardmore, J. A. 1984b. An electrophoretic study of the systematic relationships of some closely related goby species (Pisces Gobiidae). *Biological Journal of the Linnean Society* 22:107-124.
- Weir, B. S. 1996. Genetic analysis II: methods for discrete population genetic data. Pages 385-405 *in* D. M. Hillis, C. Moritz, and B. K. Mable, editors. *Molecular systematics*, 2nd edition. Sinauer Assoc., Inc., Sunderland, Massachusetts.
- Williamson, M. 1996. *Biological invasions*. Chapman and Hall, New York.





APPENDIX. Extended.

5 5 5 6  
 9 9 9 0  
 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5  
*N. melanostomus* T - C A A C A A A A A A T G C C T G A T A A T A G A A A T T

T T C A A C A A A A C T G C C T A A T A A T A G A A A T T  
 S V S

6  
 2 2 2 3  
 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4  
*N. melanostomus* A C C C A A T T A A A T A A C C T C A C C A A G T T A

A C C C A A C T A A G T A A C C T C A C C A A G T T A  
 S S V

6  
 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  
 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3  
*N. melanostomus* A G A C C T G G C C C A G A T T T A T T A A C C A T A

A G A C C T G G C C C A G A T T T A T T A A T C A T A  
 S S V







TABLE 2. Haplotype data for intraspecific variation in the mtDNA control region of *Neogobius melanostomus*. Variable nucleotide positions are numbered (see Appendix). *N* = the frequency of each haplotype.

Haplotype	<i>N</i>	Nucleotide Position										Repeats							Distribution frequency								
		5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	1	2	3	4	Avon Point	Chagrin River	Shiawassee River	Flint River	Lake St. Clair	Poland
aa	10	C	-	T	A	T	C	A	G	T	C	A	A	F	I	1	8	0	0	0	0	0	0	0	1	0	0
bb	3	.	.	.	.	.	.	.	.	.	.	A	E	F	I	1	2	0	0	0	0	0	0	0	0	0	0
cc	8	.	.	.	.	.	.	.	.	.	.	A	F	F	I	0	0	0	0	0	0	0	5	3	0	0	0
dd	10	.	.	.	.	.	.	.	.	.	G	.	G	G	G	I	2	0	0	0	0	5	0	3	0	0	0
ee	7	.	.	.	.	.	.	.	.	.	G	.	A	A	G	H	6	0	0	0	0	0	0	1	0	0	0
ff	4	.	.	.	.	.	.	.	.	.	G	.	A	A	A	H	0	0	0	0	0	0	0	2	0	0	2
gg	8	.	.	.	.	.	.	.	.	.	G	T	A	A	A	H	0	0	0	0	0	0	0	0	0	7	1
hh	3	T	T	.	.	.	.	.	.	T	A	G	T	D	A	A	H	0	0	0	0	0	0	0	0	3	0
ii	2	.	.	.	.	.	.	.	.	G	.	T	.	A	A	A	H	0	0	0	0	0	0	0	0	0	2
jj	2	.	.	.	.	.	.	.	.	T	.	G	T	A	A	A	H	0	0	0	0	0	0	0	0	0	2
kk	1	.	.	.	.	.	.	.	.	C	.	G	T	C	C	C	H	0	0	0	0	0	0	0	0	1	0
ll	1	.	.	.	.	.	.	.	.	C	.	G	.	B	B	B	K	0	0	0	0	0	0	0	0	0	1
mm	1	T	-	C	.	.	.	.	.	C	.	T	A	G	T	A	A	A	H	0	0	0	0	0	0	0	1
nn	1	.	.	.	.	.	.	.	.	.	T	.	G	T	A	A	A	J	0	0	0	0	0	0	0	0	1
oo	1	.	.	.	.	.	.	.	.	.	T	.	G	T	A	A	A	H	0	0	0	0	0	0	0	0	1
pp	1	.	.	.	.	.	.	.	.	.	T	A	G	T	A	A	A	H	0	0	0	0	0	0	0	0	1
qq	1	.	.	.	.	.	.	.	.	.	G	.	A	G	G	H	0	0	0	0	0	0	0	0	0	0	1
Total	64											10	10	5	5	10	10	10	10	10	10	10	5	5	10	11	13

TABLE 3. Nucleotide polymorphism ( $p_n$ ), haplotype diversity ( $h \pm SE$ ), and nucleotide diversity ( $\Pi \pm SE$ ) within sampling sites of *Neogobius melanostomus* (REAP version 2.0, McElroy et al. 1997) based on the entire data set.

Sampling site	$p_n$	$h \pm SE$	$\Pi \pm SE$
Avon Point	0.018	$0.644 \pm 0.152$	$0.007 \pm 0.005$
Chagrin R.	0.002	$0.356 \pm 0.159$	$0.001 \pm 0.001$
Shiawassee R.	0.000	$0.000 \pm 0.000$	$0.000 \pm 0.000$
Flint R.	0.000	$0.000 \pm 0.000$	$0.000 \pm 0.000$
Lake St. Clair	0.025	$0.844 \pm 0.080$	$0.012 \pm 0.007$
Poland	0.014	$0.564 \pm 0.134$	$0.004 \pm 0.003$
Black Sea	0.046	$0.962 \pm 0.041$	$0.009 \pm 0.004$

TABLE 4. (A) Table of combined haplotypes, using sequence information from transversions and insertions only in the left domain of the mtDNA control region for the round goby. The repeat information using transversions and insertions was re-coded as A= A through G; H= H, I, and K; and J= J. (B) All repeat information was condensed, using information from the four transversions occurring as a group (denoted as II in Table 1), all transversion that distinguished a group (marked with + in Table 1), and excluding all single transitions and transversions. Repeat information was then re-coded as A= A, B, C, D, and E; F= F and G; H= H, I, J, and K.

(A)	Base Position				Repeats				Distribution frequency						
	5	6	6	7	1	2	3	4	Avon Point	Chagrin River	Shiawassee River	Flint River	Lake St. Clair	Poland	Black Sea
Haplotype	N	21	34	3	5	1	1	4	2	10	0	5	4	0	0
		- A T	- A G	T T G	- T G	A A A	A A H	A A H	A A H	0	0	0	0	0	0
		- T G	- T G	A A A	A A H	A A H	A A H	A A H	0	0	0	0	0	0	0
		- T G	- T G	A A A	A A J	A A A	A A J	A A J	0	0	0	0	0	0	0
Total	64					10	10	5	5	10	5	5	10	11	13

(B)	Repeats				Distribution frequency							
	N	1	2	3	4	Avon Point	Chagrin River	Shiawassee River	Flint River	Lake St. Clair	Poland	Black Sea
Haplotype	N	20	8	8	0	0	0	0	0	2	0	0
		A A F H	A A F H	A A F H	0	0	0	0	5	3	0	0
		F F F H	F F F H	F F F H	2	0	5	0	0	3	0	0
		A A A H	A A A H	A A A H	0	0	0	0	0	2	11	13
Total	64				10	10	5	5	10	10	11	13

TABLE 5. Hierarchical analyses of molecular variance among haplotypes of *N. melanostomus* (AMOVA; Excoffier et al. 1992). \* = significant at  $P \leq 0.05$ . (A) Values using restricted haplotypes of entire sequence data (see Methods and Table 4A). (B) Values using restricted repeat haplotypes (see Methods and Table 4B).

(A) Combined entire data set					
Variance component		Variance	% Total variance	$\Phi$ statistic	$P$
Between regions	$\sigma_a^2$	0.080	20.85%	$\Phi_{ct} = 0.209$	0.047*
Among sampling sites	$\sigma_b^2$	0.083	21.64%	$\Phi_{sc} = 0.273$	<0.001*
Within sampling sites	$\sigma_c^2$	0.220	57.50%	$\Phi_{st} = 0.425$	<0.001*

(B) Combined repeat haplotypes					
Variance component		Variance	% Total variance	$\Phi$ statistic	$P$
Between regions	$\sigma_a^2$	0.501	57.16%	$\Phi_{ct} = 0.572$	<0.001*
Among sampling sites	$\sigma_b^2$	0.213	24.25%	$\Phi_{sc} = 0.566$	<0.001*
Within sampling sites	$\sigma_c^2$	0.163	18.60%	$\Phi_{st} = 0.814$	<0.001*

TABLE 6. Modified chi-square values from Monte Carlo tests (Roff and Bentzen 1989) for the distribution of mtDNA control region sequence haplotypes of *N. melanostomus* (REAP version 2.0, McElroy et al. 1992). \* = significant at  $P \leq 0.05$ . (A) Values using restricted haplotypes of entire sequence data (see Methods and Table 4A). (B) Values using restricted repeat haplotypes (see Methods and Table 4B).

(A) Combined entire data set		
	$\chi^2$	$P$
Between regions (North America and Eurasia)	28.2	<0.001*
Among five sampling sites (two Eurasian populations and L. Erie, Huron, and St. Clair)	53.0	<0.001*
Among L. Erie, Huron, and St. Clair	1.1	0.51
(B) Combined repeat haplotypes		
	$\chi^2$	$P$
Between regions (North America and Eurasia)	56.12	<0.001*
Among five sampling sites (two Eurasian populations and L. Erie, Huron, and St. Clair)	99.4	<0.001*
Among L. Erie, Huron, and St. Clair	32.6	<0.001*

TABLE 7. Modified chi-square values from Monte Carlo tests (Roff and Bentzen 1989) for the distribution of mtDNA control region sequence haplotypes of *N. melanostomus* between pairs of sample sites, including Lakes Erie, St. Clair, and Huron (REAP version 2.0, McElroy et al. 1992). \*\* = significant at  $P \leq 0.017$ , using a Bonferroni correction for multiple post-hoc tests (Cooper 1968; Fry et al. 1993). Above the diagonal are values using restricted haplotypes of entire sequence data (see Methods and Table 4A). Below the diagonal are values using the restricted repeat haplotypes (see Methods and Table 4B).

	Lake Erie (N = 20)	Lake St. Clair (N = 10)	Lake Huron (N = 10)
Lake Erie	-----	$\chi^2 = 1.1$ $P = 0.257$	$\chi^2 = 0.3$ $P = 0.442$
Lake St. Clair	$\chi^2 = 16.5$ $P = 0.001^{**}$	-----	$\chi^2 = 0.2$ $P = 0.375$
Lake Huron	$\chi^2 = 23.6$ $P = 0.001^{**}$	$\chi^2 = 5.0$ $P = 0.090$	-----

## Figure Legends

FIGURE 1. Drawing of (A) *Neogobius melanostomus* and (B) *Proterorhinus marmoratus* (adapted from Miller et al. 1986, with permission).

FIGURE 2. North American collection sites for *Neogobius melanostomus* are 1= Shiawassee R., near Linden, Michigan, 2= Flint R., near Russelville, Michigan, 3= Lake St. Clair, 5= off Avon Point, and 6= off the Chagrin R. *Proterorhinus marmoratus* was collected at sites 3= Lake St. Clair and 4= St. Clair River. (Distribution information from Biological Resources Division, USGS, 1997).

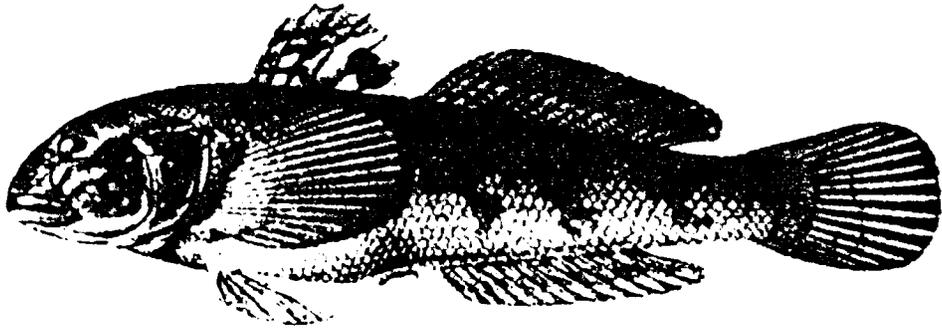
FIGURE 3. European collection sites for *Neogobius melanostomus* are numbered 7-8. 7= Gulf of Gdansk, Poland, and 8= Black Sea, near Varna, Bulgaria. The Ponto-Caspian region includes the Marmara, Black, Azov, Caspian, and Aral Seas.

FIGURE 4. A schematic diagram of the left domain of the control region for (A) *Neogobius melanostomus* and (B) *Proterorhinus marmoratus*. Features such as the repeats, short repeat (SR), termination associated sequences (TAS) and the central conserved section (CCS) are indicated. (A) | = intraspecific substitutions, (B) | = interspecific substitutions.

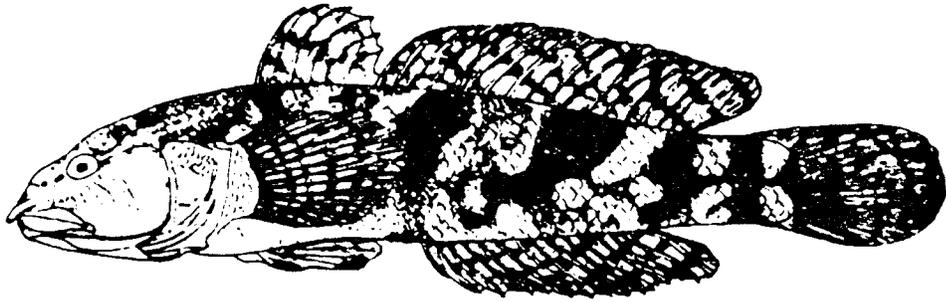
FIGURE 5. Neighbor-joining tree for haplotypes of *Neogobius melanostomus* (based on the entire data set; Table 2) using Kimura (1980) two-parameter distances in the computer program MEGA (Kumar et al. 1993). The *Proterorhinus marmoratus* haplotype was used as an outgroup. Distances may be calculated by adding horizontal branch lengths between pairs. Percentages

indicate bootstrap support for nodes computed from 1,000 replicates. Major types present in the Great Lakes are labeled 1-4, and major groups are denoted as I and II (see Results).

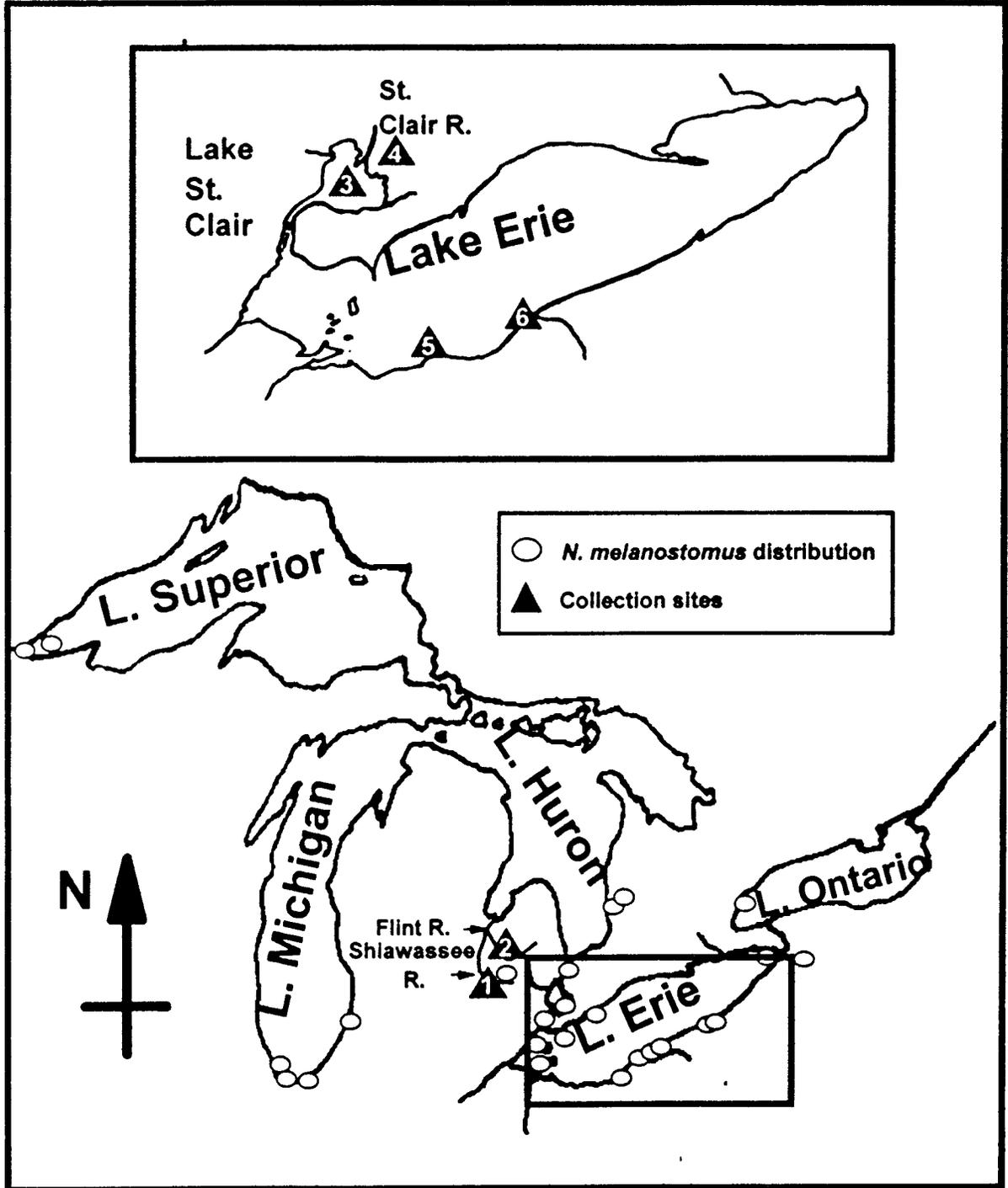
FIGURE 6. 50% majority rule consensus tree (Margush and McMorris 1981) of the 12 most parsimonious trees for haplotypes of *Neogobius melanostomus* (based on the entire data set; Table 2) using PAUP\* (Swofford 1998) and a branch-and-bound search (Hendy and Penny 1982). *Proterorhinus marmoratus* was used as the outgroup. Percentages indicate consensus support/bootstrap support for nodes computed from 1,000 replicates. Major types present in the Great Lakes are labeled 1-4 (see Results).



**(A)**

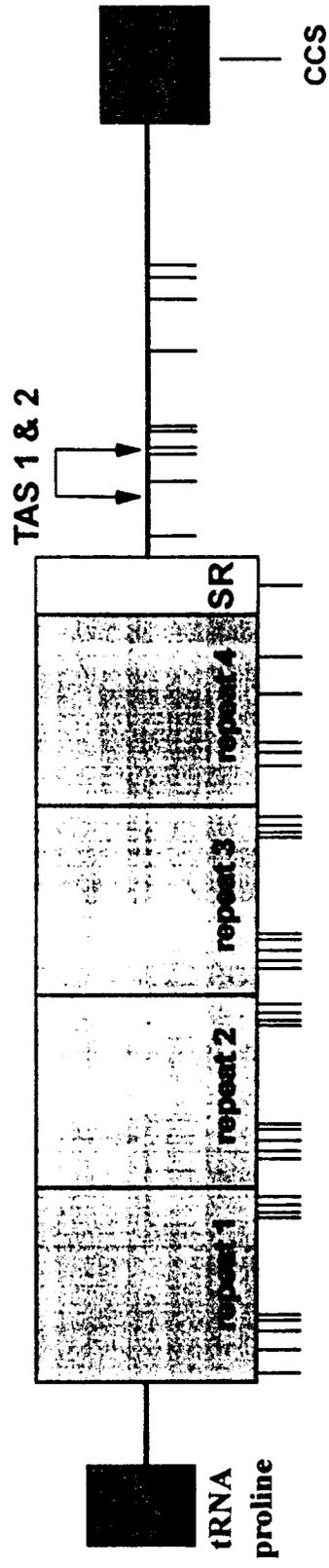


**(B)**

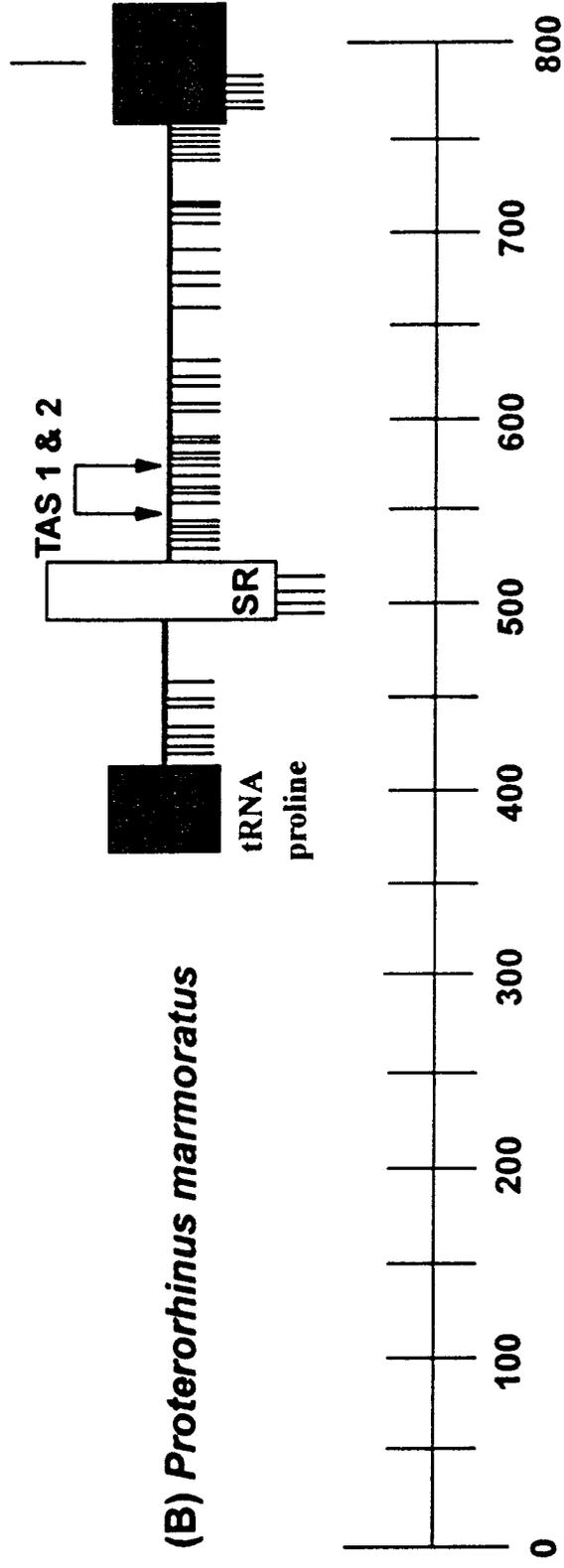




(A) *Neogobius melanostomus*

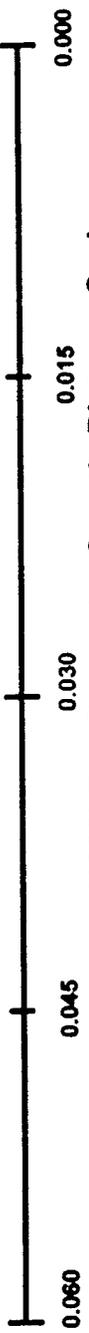


(B) *Proterorhinus marmoratus*





*P. marmoratus*



Kimura (1980) Two Parameter Genetic Distance Scale

