Structure and Affinities of Freshwater Sea Lamprey (*Petromyzon marinus*) Populations¹

PETER F. BRUSSARD, MARJORIE COLLINGS HALL, AND JANET WRIGHT

*Section of Ecology and Systematics, Cornell University, Ithaca, NY 14850, USA*


We have examined genetic differentiation at 23 loci in 12 populations of sea lamprey, *Petromyzon marinus*, by means of starch gel electrophoresis. Based on two measures of overall genetic distance and two clustering methods, our analysis shows that there are three genetically distinct groups of lamprey in eastern North America: (1) anadromous populations plus those in Lake Champlain, (2) populations in Lake Erie and the upper Great Lakes, and (3) populations in Lake Ontario and three interior New York Lakes (Cayuga, Seneca, and Oneida). Analysis of population subdivision using contiguity partitions and simultaneous test procedures (STP) confirms the above conclusions and offers additional insight into the genetic structure of lamprey in this area.

*Key words:* sea lamprey, *Petromyzon marinus*; genetic variation, electrophoresis, population structuring


Received October 9, 1980
Accepted June 23, 1981

It is usually assumed that sea lamprey (*Petromyzon marinus*) populations in the upper Great Lakes are recently derived from Lake Ontario stocks, having gained access to those waters with the opening of the Welland Canal in 1829 (Anon. 1967; Scott and Crossman 1973). However, the origin of lamprey populations in Lake Ontario, Lake Champlain, and several of the smaller interior New York lakes (Seneca, Cayuga, and Oneida) is unclear. Some or all of these may be relict populations, landlocked since early postglacial times, or they may have gained access to these waters in historical times via various branches of the New York State Barge Canal System. This matter is of more than academic interest for at least two reasons. First, if lamprey are indeed endemic, it is possible that prey species from these waters may have evolved some degree of resistance to lamprey predation. Such resistance would be useful in stocks used to repopulate areas where lamprey predation is a problem. Second, if lamprey really are endemic to certain lakes, managers may also want to reexamine a control philosophy aimed at exterminating a native species which has coexisted with the other species in these waters for over 10000 years.

Potential postglacial ancestral sources of lamprey in Ontario and the interior New York lakes were (1) the Delaware–Susquehanna drainages during the time that glacial Lake Watkins overflowed to the south via Seneca Lake, (2) one or more of the later stages in the recession of the ice front with the major drainage outlet through the Hudson–Mohawk system, and (3) the St. Lawrence River, especially during the marine incursion of the Champlain Sea. The second two possibilities would also account for the presence of lamprey in Lake Champlain.

In contrast, the Hudson River is the most likely source of recent invasion via the Erie Canal, a branch of which also

¹This paper forms part of the Proceedings of the Stock Concept International Symposium (STOCS) convened at Alliston, Ontario, September 29 — October 9, 1980.
connected with Lake Champlain. (Completion of the Chemung branch in 1833, joining the Susquehanna River and Seneca Lake, seems too recent to account for definite presence of lamprey in Lake Ontario in 1835.)

The major objective of this project was to undertake a comprehensive electrophoretic study to determine the genetic affinities of lamprey populations in waters of the northeastern United States and Canada. These would include samples from possible endemic stocks (Lake Ontario, Lake Champlain, and Cayuga, Seneca, and Oneida lakes), recently colonized stocks (Lakes Erie, Huron, Michigan, and Superior), and potential ancestral marine stocks, i.e. anadromous populations from the St. Lawrence, Susquehanna, Delaware, and Hudson rivers. Fulfillment of this goal would give insight into the current population structure of P. marinus in these areas and might provide critical evidence for determining the origins of the freshwater populations.

Population Structure

The amount of genetic differentiation among populations of a species, which results primarily from inbreeding and isolation, is referred to as that species' population structure (Selander 1975; Selander and Kaufman 1975). Despite the fact that information on population structure is critical to proper resource management, methods have not been available until recently for generating such data. Attempts have been made using morphological variants, but these are usually affected by an unknown number of genes and an unknown environmental component, making inferences on population structure based on such characters tentative at best. However, recently developed electrophoretic techniques for demonstrating protein variation permit the description of genotypic differences at large numbers of structural gene loci, and it is now possible to describe the population structure of almost any species in detail.

Electrophoretic variants (electromorphs) reflect simple genetic differences, and their phenotypes are codominantly expressed as bands on gels. When available, breeding data almost invariably confirm the presumed genetic model for these markers, as progeny resulting from crosses conform to the Mendelian proportions expected from the parental phenotypes. The assumption of genetic variation in species where breeding experiments are not possible is validated by criteria such as repeatability, similarity to known genetic variants in other species, and conformance of phenotypic frequencies to Hardy–Weinberg expectation.

Extensive studies of protein variation have shown spatial variation in gene frequencies among populations of most species. Animals with good dispersal abilities, such as Drosophila, have very similar gene frequencies throughout their ranges (e.g. Ayala et al. 1974), while species with limited mobility or pronounced homing tendencies manifest obvious geographic differentiation (e.g. Selander and Kaufman 1975). In many cases, population samples can be consistently distinguished on the basis of phenotype frequency differences or by the presence of unique variants, and such techniques have been used with considerable success in the genetic characterization of fish stocks. For example, trout and salmon populations in the Pacific Northwest have been investigated intensively using electrophoretic techniques by workers at the Northwest Fisheries Center (reviewed by Utter et al. 1976). Franzin and Clayton (1977) were able to determine postglacial routes of gene flow among western Canadian lake whitefish (Coregonus clupeaformis) populations by using electrophoretic markers, and these authors then related the biochemical data to the glacial refugia in which the whitefish are believed to have survived the Wisconsin glaciation. Numerous other examples of the application of biochemical techniques to biogeographical problems can be found in the recent literature.

The success of these methods for determining population affinities results from the fact that electrophoretically detectable genetic differentiation begins to accumulate in subdivided populations as soon as their gene pools become isolated. These differences result from a combination of various stochastic (founder effects, "bottlenecks," and drift) and deterministic (selective) factors. In the absence of gene flow, the rates of divergence in electrophoretic characters seem to be reasonably time-dependent (Wilson et al. 1977), and genetic distances among well-isolated contemporary populations, calculated from electrophoretic data, tend to reflect the histories of the populations. These histories not only include patterns of derivation from common stocks but may also reflect the nature of colonization events and the results of low population size.

Methods and Materials

Ammocoetes were obtained by electroshocking 11 localities from April to September, 1979, and June to August, 1980. In addition, transformers and adult lampreys were obtained from the fish weir at Cayuga Inlet, Cayuga Lake, New York, in April, 1979, and adult lampreys were obtained from the Connecticut River in June, 1980. Specimens were transported live to the laboratory and then immediately frozen at −80°C.

Protein variation was studied in both muscle and liver tissues. In preparation for electrophoresis the head, gut, skin, and notochord were removed. Animals less than 50 mm in length were ground whole. Liver and muscle were minced in equivalent volumes of grinding buffer (0.01 M Tris, 0.001 M EDTA, 5 × 10⁻⁶ M NADP, pH adjusted to 7.0 with HCl) and allowed to soak for 30 min. Following centrifugation for 20 min at 2500 g at 0–5°C, the clear supernatant was removed. Tissue extracts were stored at −80°C and most samples were electrophoresed within 2–3 wk.

The electrophoretic apparatus, gel and electrode buffers, and staining procedures used in this study were similar to those described by Selander et al. (1971), with the following exceptions: three other buffer systems were also used, "Tris-Whitt" (Whitt 1970), "ASAC" (Avise et al. 1975), and "OM" (O'Brien and Mackintrye 1969). Because of a recent change in the properties of Electrostarch, we found it necessary to mix it with Sigma starch (Sigma Chemical Company, St. Louis, Missouri 63178) for satisfactory results. For gels made with Selander's buffer No. 2, 30 g Electrostarch was combined with 17 g Sigma starch; all other systems worked well with 37 g Electrostarch: 10 g Sigma starch.

Initially a survey was conducted to determine which enzyme loci could be resolved with our techniques. Twenty-eight different enzyme assays were tested on all 12 buffer
systems (Selander's 1–9, Tris–Whitt, ASAC, and OM). A side-by-side comparison of ammocoetes, transformers, and adults from the same collecting site was carried out to determine if any ontogenetic differences exist.

Genotype frequencies were obtained by direct count from the electrophoretic phenotypes on the gels. Expected genotype frequencies under Hardy–Weinberg equilibrium were calculated using Levene's (1949) formula, and these were compared with observed frequencies by a $\chi^2$-test for goodness of fit. Measures of between-population genetic similarity and distance were calculated using Rogers' (1972) and Nei's (1972) methods. To visualize the genetic affinities of the 12 populations, these distance measures were used to construct UPGMA phenograms (Sneath and Sokal 1973) and minimum length Wagner trees (Farris 1972). Geographical partitioning of the populations into internally homogeneous sets was done by modifications of a procedure suggested by Gabriel and Sokal (1969) and Pielou (1979). The entire area from which the samples had been obtained was subdivided into geographically connected subareas, based on water connections past and present. This information was obtained primarily from Whitford (1906) and USDA (1970). Then, the 12 samples were combined into internally homogeneous subpopulations based on the homogeneity of the observed allelic frequency distributions (STP procedure; Sokal and Rohlf 1969, p. 582).

Results

Our current understanding of protein variation in Petromyzon marinus is based on 775 ammocoetes collected from 11 localities as follows: 127 from Cayuga Inlet, Cayuga Lake; 65 from Fish Creek, Oneida Lake; 114 from Little Sandy Creek, Lake Ontario; 36 from Oquaga Creek, Delaware River; 71 from Bouchouxville Creek, Delaware River; 106 from Putnam Creek, Lake Champlain; 53 from Catherine Creek, Seneca Lake; 31 from Roelli Jensen Kill, Hudson River; 30 from Nebagamon River, Lake Superior; 38 from Crooked Creek, Lake Erie; 104 from Cattaraugus Creek, Lake Erie; and from 60 transformers and 46 adult lampreys from Cayuga Inlet, Cayuga Lake, and 64 adult lampreys from the Connecticut River.

We have now analyzed variation in 18 proteins, presumably encoded by 23 loci, in all of these individuals. In the following account loci are numbered and alleles designated in order of their decreasing anodal mobility. The loci examined and the buffer systems used are esterase (EST) and general protein (GP), both on Tris-hydrochloric acid buffer (Selander et al. 1971, buffer 1); alcohol dehydrogenase (ADH), peroxidase, two loci (PER-1, PER-2), and phosphohexose isomerase, two loci (PHI-1, PHI-2), with all discontinuous Tris-citrate buffer (buffer 3); xanthine dehydrogenase (XDH), α-glycerophosphate dehydrogenase (α-GPDH), and phosphoglucomutase (PGM-2), all with continuous Tris-citrate buffer (buffer 5); glutamate dehydrogenase (GDH) with Tris-versene-borate buffer (buffer 6); lactate dehydrogenase (LDH), and octanol dehydrogenase (ODH), both with phosphate buffer (buffer 7); glyceraldehyde-3-phosphate dehydrogenase (G-3-PD) and malic enzyme (ME), both with Tris-malate buffer (buffer 9); malate dehydrogenase, two loci (MDH-1, MDH-2), glucose-6-phosphate dehydrogenase (G-6-PD), 6-phosphogluconic dehydrogenase (6-PGD), and phosphoglucomutase (PGM-1), all with “Tris–Whitt” (Whitt 1970); and isocitrate dehydrogenase, two loci (IDH-1, IDH-2), and glutamate oxaloacetate transaminase (GOT), all on ASAC (Avise et al. 1975).

Of the preceding loci, five (α-GPDH, PHI-1, PGM-1, PGM-2, and MDH-2) proved to be polymorphic in most populations. IDH-1 was also found to be slightly polymorphic, as evidenced by several heterozygote phenotypes from the Delaware, Hudson, and Connecticut River (anadromous) populations. MDH-2 and α-GPDH are structural dimers; the heterozygotes show three bands. PGM-1 and PGM-2 are monomeric with two-banded heterozygotes. PHI and IDH homozygotes are normally expressed as three-banded phenotypes with the most anodal and most cathodal band representing homodimeric products of two loci. A band of intermediate mobility represents the heterodimer. Allelic variants at either locus result in multi-banded phenotypes representing the various possible homo- and hetero-dimeric combinations. These electrophoretic patterns are consistent with their probable homologues in other fishes.

Comparisons of ammocoetes, transformers, and adults revealed no ontogenetic changes except for slight mobility shifts between ammocoetes versus transformers and adults at both the PER-1 and PER-2 loci. It is presently unknown whether these changes result from different loci being active at different stages of the life cycle or from posttranslational changes in the proteins themselves, although the latter explanation seems most likely.

α-GPDH had two common alleles, α-GPDH$^a$ and α-GPDH$^b$, both of which were present in all the samples. PHI-1$^b$ was the common allele at that locus, while PHI-1$^a$ occurred in low frequencies in most populations except for Oneida Lake where it reached a frequency of 0.32. It was missing in our sample from Lake Superior. PHI-1$^b$ also occurred in low frequencies in most populations, but was missing from Seneca Lake and Lake Champlain. PHI-1$^a$ was observed once as a heterozygote in the Connecticut River sample. Both PGM-1$^a$ and PGM-1$^b$ were present in varying frequency in all populations examined. PGM-2$^a$ was missing from Oquaga Creek as well as from Seneca Lake, but has reached a relatively high frequency of 0.25 in Lake Superior and Lake Erie. PGM-2$^b$ is a rare variant found only in the anadromous samples. MDH-2$^a$ was fixed in Seneca Lake and was the most common allele in all other samples. MDH-2$^b$, when present, varied in frequency from 0.01 to 0.27. These data are summarized in Table 1.

The remaining 17 loci were monomorphic and showed no evidence of intra- or inter-population variability. Overall heterozygosity of P. marinus averaged about 6%, a value similar to that reported for most other fishes (e.g. Avise and Ayala 1976).

Four of 58 comparisons of observed and expected genotypic proportions deviated from Hardy–Weinberg equilibrium, one at the 0.01 significance level and three at the 0.05 level. In each case the deviation was accompanied by a substantial deficiency of heterozygotes. The most reasonable explanation for this is probably a Wahlund effect from combining year-classes, as all ammocoetes collected at a site (ranging from perhaps 2–6 or more years old) were pooled for our sample.

We used two measures of genetic distance and two
TABLE 1. Allele frequencies and sample sizes \( (N) \) at the polymorphic loci for 12 populations of \( P. \) marinus.*

<table>
<thead>
<tr>
<th></th>
<th>CAY</th>
<th>ONI</th>
<th>ONT</th>
<th>SEN</th>
<th>BXC</th>
<th>OQC</th>
<th>CHA</th>
<th>HUD</th>
<th>SUP</th>
<th>CRC</th>
<th>CAC</th>
<th>CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPDH</td>
<td>0.58</td>
<td>0.61</td>
<td>0.54</td>
<td>0.64</td>
<td>0.40</td>
<td>0.32</td>
<td>0.27</td>
<td>0.29</td>
<td>0.25</td>
<td>0.43</td>
<td>0.45</td>
<td>0.32</td>
</tr>
<tr>
<td>GPDH</td>
<td>0.42</td>
<td>0.39</td>
<td>0.46</td>
<td>0.36</td>
<td>0.60</td>
<td>0.68</td>
<td>0.73</td>
<td>0.71</td>
<td>0.75</td>
<td>0.57</td>
<td>0.55</td>
<td>0.68</td>
</tr>
<tr>
<td>( N )</td>
<td>50</td>
<td>36</td>
<td>49</td>
<td>22</td>
<td>57</td>
<td>30</td>
<td>99</td>
<td>31</td>
<td>30</td>
<td>38</td>
<td>91</td>
<td>64</td>
</tr>
</tbody>
</table>

*Locality abbreviations are as follows: CAY, Cayuga Lake; ONI, Oneida Lake; ONT, Lake Ontario; SEN, Seneca Lake; BXC, Bouchouxville Creek; OQC, Oquaga Creek; CHA, Lake Champlain; HUD, Hudson River; SUP, Lake Superior; CRC, Crooked Creek; CAC, Cattaraugus Creek; CON, Connecticut River.

Simultaneous test procedures for heterogeneity of allele frequencies (STP; Sokal and Rohlf 1969, p. 582) indicated that allele frequencies at every polymorphic locus were significantly different among various groups of populations. All the anadromous samples (CON, BXC, OQC, HUD) were homogeneous at all loci, as were the Lake Erie samples (CAC, CRC). All other populations showed statistically significant frequency differences at one or more loci. The results of this analysis and the contiguity partitioning are shown in Fig. 2.
that there are three major groups of populations in eastern North America (Fig. 1). The Delaware, Connecticut, and Hudson River populations represent anadromous stocks, and these seem to show as much differentiation among tributary creeks in the Delaware River as they do between the major river systems. The Lake Champlain population is closely related to the anadromous stock. The second major cluster includes Lake Ontario and the interior New York lakes (Seneca, Cayuga, and Oneida), and the third consists of the samples from Lake Erie and Lake Superior.

Although this pattern is clearly consistent with the geography of the region, interpreting the history of these populations from these data alone is not altogether straightforward. For this kind of study the assumption is usually made that allozymes behave as essentially neutral markers of genetic continuity or discontinuity, and that the allelic composition of any derived population will reflect primarily its ancestral source and the extent of gene flow from neighboring populations. These assumptions are probably true in general, and electrophoresis has proved to be a powerful tool for analyzing both population structure and history (e.g. Brussard 1975). Nevertheless, allozyme frequencies in any population are potentially influenced by several other confounding factors. First, the type of event leading to the population’s origin can have some influence. The amount of genetic information present in a derived population will be smaller if it arose from a few colonizing individuals rather than from the isolation of a large segment of the ancestral population; furthermore, if the derived population remains small for several generations after its founding, many alleles will be lost and others will drift to abnormally high frequencies. Second, previous episodes of gene flow can have large effects on the genetic composition of contemporary populations. For neutral or nearly neutral alleles, just a trickle of gene flow can restore lost alleles and erase differences resulting from previous events. Third, local selection pressures may have some effects as well. Selection can maintain allelic differences in the face of considerable gene flow, provided that the selective pressures are strong enough. While there is little evidence to suggest that local selection pressures exert a powerful influence on most allozyme frequencies, slight deviations from expected genotypic proportions or minor frequency shifts at some loci may indicate the effects of selection (e.g. Koch and Williams 1978). Furthermore, it is even possible that genetic divergence in some electrophoretic characters may depend on a synergistic relationship between selection and founder effects to a greater degree than is generally suspected (Baverstock et al. 1979). When a small isolate splits off from a parent population it may be free to undergo a “peak shift” (Wright 1931, 1980), or even a mild “genetic revolution” (Carson 1975) resulting in a general realignment of equilibrium allele frequencies.

The genetic structure of the P. marinus populations we have examined may have been affected by a combination of several of the factors mentioned above. The genetic distances among these populations are small, and result from slight, but generally significant, differences in allelic frequencies and from the presence or absence of a few rare alleles. No locus is fixed for alternate alleles in different populations; no alleles are unique to any one population; and in most cases the common allele at a locus in one population is the common

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Population structure of P. marinus. A. Genetic continuity based upon G-tests for heterogeneity of allele frequencies at five polymorphic loci among geographically contiguous populations. Broken line means significant heterogeneity at all five loci; single line, significant heterogeneity at four loci; double line, significant heterogeneity at three loci; triple line, significant heterogeneity at two loci; quadruple line, significant heterogeneity at one locus. B. Continuity based on Rogers’ coefficient of genetic similarity using allelic frequency data from the five polymorphic loci. Broken line means similarity coefficient of 0.79–0.81; single line, 0.82–0.85; double line, 0.86–0.89; triple line, 0.90–0.93; quadruple line, 0.94 or greater. In both A and B increasing line thickness indicates increasing genetic continuity. ANA = all anadromous populations, ERI = both Lake Erie populations. All other abbreviations as in Table 1.

**Discussion**

We had two major objectives when we began this study. The first concerned the historical biogeography of sea lamprey populations in Lake Ontario, Lake Champlain, the Finger Lakes, and Oneida Lake. Specifically, we wanted to know whether lampreys entered these waters in historical times via the canal systems, or whether some or all of them were relic populations, isolated since the last glacial retreat. The second objective was to define the current structure of sea lamprey populations in the study area and to look for any evidence of gene flow between major population units or for evidence of substructuring within any of these waters.

At this point we have had only limited success with the former goal and somewhat more with the latter. Our current understanding of the genetic relationships among lamprey populations, based on patterns of variation at 23 loci, indicates
allele in all others. The average Nei distance among all P. marinus populations is 0.008; an equivalent level of differentiations has been observed by Merritt et al. (1978) in populations of dace, Rhinichthys cataractae, within the Connecticut River drainage.

Because we cannot preclude the effects of selection working on at least some loci, and because we cannot discount the possibilities of previous gene flow occurring among some of these populations, any interpretation of the history of lamprey colonization in this region based on genetic distance data must be tempered by some subjective judgments. Nevertheless, the contiguity diagrams in Fig. 2 enable us to examine these data more closely and within a geographical perspective. In this figure all contiguous populations which had no significant allele frequency differences at any loci (i.e. all anadromous samples, both Lake Erie samples) are combined. Examination of Fig. 2 leads to the following observations. First, the close relationship between the Lake Champlain population and the anadromous stock may indicate that colonization of Lake Champlain occurred as a result of postglacial incursion of the Champlain Sea, or it may indicate a recent colonization via the Champlain canal from the Hudson River. Significant gene flow between Lake Champlain populations and those in the Hudson River cannot be discounted.

Second, Lake Ontario seems to be the most likely source for the colonization of Lake Erie, rather than the Hudson River via the Erie Canal, as is occasionally suggested (T. M. Jolliff personal communication). The moderately large degree of genetic differentiation between the Ontario and Erie populations may have resulted from a rather substantial bottleneck during the colonization period.

Third, while there is certainly substantial differentiation between the anadromous populations and those in Lake Ontario, Oneida, and the Finger Lakes, the data on hand cannot unambiguously distinguish between endemcity for those populations or a more recent arrival. It is certainly possible that bottleneck effects could also have been operative in the colonization of Lake Ontario. On the other hand, positive evidence of the presence of lamprey in Lake Ontario in 1835 (Lark 1973) may provide additional evidence for endemcity, since it is unlikely that a small founder population from Hudson River stocks would have invaded, differentiated, and proliferated in the 11 years following the completion of the water connection in 1824-25. The Erie colonization poses no such time constraints, since the approximately 100 years between the completion of the Welland Canal and the recorded appearance of P. marinus in Lake Erie (Applegate 1950) should have provided ample time for considerable genetic differentiation to occur. Lake Erie is not particularly favorable lamprey habitat, and it is quite possible that a small colonizing population existed at a very low level there during the intervening century. This would have allowed the PGM-1 and PGM-2 alleles to drift to a relatively high frequency and would account for the apparent loss of the PHI-1 allele. Lake Superior is clearly genetically very similar to Lake Erie. It was the last of the upper Great Lakes to be colonized, and as such it may represent the end of a cline of differentiation in those waters.

Fourth, there is obviously substantial similarity among the Lake Ontario and interior New York populations. Whether this similarity results from colonization from a single stock, either at the close of the Pleistocene or more recently, from gene flow among these populations, or both, is unknown, although scattered historical data tend to support endemcity (D. Webster unpublished data). There may be substantial migration among these populations as well, since there are reports of lamprey in several tributary streams in the Seneca-Oneida-Oswego River systems, which connect these lakes (D. Webster personal communication). The relative distinctiveness of Lake Ontario from Oneida Lake is also worthy of mention. If Oneida Lake contributes to the Lake Ontario population by downstream drift of transforming larvae as has been suggested by some workers (Webster 1979), this contribution, in terms of gene flow at least, may be rather small, provided, of course, that our sample is typical of the lake as a whole.

These data have provided us with a good general understanding of lamprey population structure in this region. However, a resolution of the endemcy issue and a clarification of certain finer details of this pattern will depend on additional data. Samples from key areas are currently being analyzed, and these should enhance our results substantially and facilitate the interpretation of these rather intriguing data.

**Acknowledgments**

Special thanks are extended to those who assisted with the field and laboratory work, especially Cara Adler, Vicky Meretsky, and John Wilson. Several employees of the New York State Department of Environmental Conservation and the Great Lakes Fisheries Commission helped us obtain specimens, and Charles Krueger furnished the Lake Superior sample. Thomas M. Jolliff and Steven Gourley offered valuable comments on an earlier draft of this manuscript.

This study was generously supported by a grant from the Great Lakes Fisheries Commission.

**References**


