Fisheries and Oceans Canada
Science

Pêches et Océans Canada
Sciences

C S A S

Canadian Science Advisory Secretariat

S C C S

Secrétariat canadien de consultation scientifique

Research Document 2011/008

Maritimes Region

Document de recherche 2011/008

Région des Maritimes

Genetic and Phenotypic Diversity and Divergence in Sympatric Lake Utopia Rainbow Smelt (*Osmerus mordax*)

Diversité et divergence génétiques et phénotypiques chez les populations sympatriques d'éperlan arc-en-ciel du lac Utopia (*Osmerus mordax*)

I.R. Bradbury[1]*, R. Bradford[2], and P. Bentzen[1]

[1]Marine Gene Probe Laboratory
Biology Department
Life Sciences Centre
Dalhousie University
Halifax, Nova Scotia, Canada

[2] Population Ecology Division
Bedford Institute of Oceanography
PO Box 1006
Dartmouth, NS, Canada

*Author to whom correspondence should be addressed
email: ibradbur@me.com

ISSN 1499-3848 (Printed / Imprimé)
ISSN 1919-5044 (Online / En ligne)

Canada

TABLE OF CONTENTS

Correct citation for this publication:
Le présente publication doit être citée comme suit :

Bradbury, I.R., R. Bradford, P. Bentzen. 2011. Genetic and Phenotypic Diversity and Divergence in Sympatric Lake Utopia Rainbow Smelt (*Osmerus mordax*). DFO Can. Sci. Advis. Sec. Res. Doc. 2011/008: vi + 28p.

ABSTRACT

The Rainbow Smelt complex in Lake Utopia New Brunswick represents an independent occurrence of multiple sympatric ecological forms, the identification and status of which remains unresolved. Here we re-evaluate the presence of phenotypic and genetic divergence within Lake Utopia smelt using samples of spawning individuals collected in 1980-1981, 1990-1991, 2002-2003, and 2010. Phenotypic divergence was examined using several meristic traits (gill raker, and fin ray counts) and linear morphometric measurements. Gill raker counts generally support the presence of two groups of smelt, whereas variation in body length reveals three distinct modes. Principle Component Analysis (PCA) and Quadratic Discriminant Function (QDFA) analysis of linear morphometrics support hypotheses of two or three size-based groups (98-99% accurate assignment) and the statistical removal of size significantly reduced (~65%) assignment success. Data on 10 microsatellite loci spanning 20 years consistently support a hypothesis of two discrete populations. Principle Coordinate Analysis and Bayesian clustering using STRUCTURE revealed the presence of two populations, one associated with the two smaller body length modes, and the other consisting of individuals greater than ~170 mm fork length. Evidence of hybridization between the forms was observed in both morphs. Moreover, multiple small-form individuals were identified with 'large form' genotypes. Overall, the results support the presence of two genetically distinct forms in Lake Utopia, with evidence of phenotypic divergence within the small form. The recurring presence of hybrid individuals and gene flow from the large to small forms suggests that the stability of the pair will likely depend on the persistence of suitable local environmental conditions, and presents the possibility of introgressive meltdown of the forms if conditions change. Further study of the rate of hybridization, the fate of hybrids, and the selective landscape associated within the lake are needed to evaluate the status and allow continued conservation of smelt in Lake Utopia.

RÉSUMÉ

Le complexe que forment les éperlans arc-en-ciel du lac Utopia, au Nouveau-Brunswick, représente une occurrence indépendante de multiples formes écologiques sympatriques de ce poisson, dont l'identification et la situation restent à élucider. Nous réévaluons ici l'existence d'une divergence phénotypique et génétique chez l'éperlan arc-en-ciel du lac Utopia en nous fondant sur des échantillons de géniteurs prélevés en 1980-1981, 1990-1991, 2002-2003 et 2010. La divergence phénotypique a été étudiée d'après diverses caractéristiques méristiques (dénombrement des branchitechnies et des rayons de nageoire) et mesures morphométriques linéaires. Il ressort de manière générale du dénombrement des branchitechnies que deux groupes d'éperlans seraient présents, tandis que la variation de la longueur corporelle reflète trois modes distincts. L'analyse des principales composantes et l'analyse discriminante quadratique des valeurs morphométriques linéaires soutiennent les hypothèses de présence de deux ou trois groupes de tailles (98-99 % d'exactitude dans l'assignation); la suppression statistique de la taille diminuait notablement (~65 %) le succès d'assignation. Les données sur 20 ans concernant 10 locus microsatellites soutiennent avec constance l'hypothèse de deux populations distinctes. L'analyse des principales composantes et le regroupement bayésien fondé sur STUCTURE révélaient la présence de deux populations, l'une associée aux deux plus petits modes de taille et l'autre composée d'individus de ~170 mm de longueur à la fourche. Les signes d'une hybridation entre les formes apparaissaient chez les deux morphes. De plus, de multiples individus de la petite forme de l'espèce ont été identifiés avec des génotypes de sa grande forme. Dans l'ensemble, les résultats laissent croire à l'existence d'éperlans de deux formes génétiquement distinctes dans le lac Utopia, avec des signes de divergence phénotypique chez la petite forme. La présence récurrente d'individus hybrides et le flux génétique de la grande forme vers la petite donnent à penser que la stabilité de la paire va probablement dépendre de la persistance de conditions environnementales locales propices et qu'une fusion introgressive des formes est possible si les conditions changent. De plus amples études du taux d'hybridation, du devenir des hybrides et de l'écopaysage sélectif associé au lac sont nécessaires pour évaluer l'état des populations d'éperlan du lac Utopia et continuer d'œuvrer à leur conservation.

INTRODUCTION

Adaptive radiations represent a significant source of diversity in temperate ecosystems where the colonization of new habitats has occurred following Pleistocene glaciation in (e.g., Taylor and Bentzen 1993a; Bernatchez et al. 1999). These radiations are often associated with the evolution of alternate ecological strategies (Claytor and Verspoor 1991; Schluter & McPhail 1992; McPhail 1994; Lu and Bernatchez 1998; Rogers and Bernatchez 2006) and result in morphologically bimodal and phenotypic divergent populations (e.g., Bell 1976, 1989 Taylor and Bentzen 1993a). However, given the potential for contact and continued gene flow, these bimodal radiations often appear fragile and susceptible to collapse and introgression (e.g., Hendry et al. 2006; Taylor et al. 2006). The continued management and conservation of these bimodal pairs and ecological species requires a clear understanding of the diversity present and the selective forces driving divergence and stability.

The rainbow smelt, *Osmerus mordax* (Mitchill), is a small fish found in coastal and freshwater systems throughout northeastern North America from New Jersey to Labrador (Nellbring 1989). Smelt display high levels of phenotypic diversity, life history variation, and habitat types (Nellbring 1989). Freshwater microphageous small forms are common throughout the range and are usually characterized by a smaller size (60-150 mm), relatively larger eyes, and greater number of gill rakers (Kendall 1927; Copeman 1977; Copeman and McAllister 1978), contrasting the large macrophageous form which typically matures at a larger size (150-300 mm). These forms co-occur as sympatric pairs in several locations throughout eastern North America and each occurrence seems to be the result of an independent ecological divergence (Taylor and Bentzen 1993b).

Early work on smelt phenotypic variation debated the genetic basis of the large and small forms and the taxonomic status of the small morph (Green 1939; Rupp and Redmond 1966; Copeman and McAllister 1978). Transplant experiments during the 1960's provided conflicting results on the nature of this divergence; first concluding a largely environmental influence (Rupp 1959; Rupp and Redmond 1966), and later concluding a genetic basis (Copeman and McAllister 1978). Based on these studies the small morph was subsequently given species status (i.e. *Osmerus spectrum*, Lanteigne and McAllister 1983), which was later challenged using molecular genetic analysis documenting the independent sympatric origins of each occurrence (Taylor and Bentzen 1993b). Recently, the presence of a third form of intermediate-size smelt has been reported in Lake Utopia, New Brunswick (Curry et al. 2004). The possibility of an additional form, its relationship to the previously described morphs, and the overall stability of the sympatric pair in Lake Utopia requires further evaluation if this complex occurrence is to be understood and managed successfully.

Here we re-evaluate the presence of phenotypic and genetic divergence within Lake Utopia smelt using samples of spawning individuals collected over two decades to quantify the number of discrete populations present, and the potential for hybridization or fusion among the forms present. Some of the morphological data used in this study have been published previously in modified form (e.g. Lanteigne and McAllister 1983, Taylor and Bentzen 1993 a,b); however we bring together all of these data in the context of a comprehensive genetic analysis with the objective of providing new insight into evolutionary ecological dynamics of rainbow smelt in Lake Utopia.

METHODS

Lake Utopia is located in southwestern New Brunswick and has a length of 7.2 km and a mean depth of 11m (COSEWIC 2008, Figure 1) Mature adult smelt were collected by dip and fyke netting during the spawning and post-spawning period in 2002, 2003, and 2010 from five previously identified (Curry et al. 2004) spawning locations (see Figure 1, Table 1). Fin clips were taken and immediately placed in 95% ethanol and fish were immediately frozen for later otolith extraction and morphological examination. In short, individuals were collected from five locations, with small morph being observed in the brooks (e.g., Unnamed Brook, Scout Brook, and Mill Brook), and the large morph in the outlet streams for connected lakes (e.g., Trout Lake Stream and Mill Lake Stream). Material from previous collections was also used including tissue and phenotypic data from Taylor and Bentzen (1993a) and phenotypic data from Lanteigne and McAllister (1983) from the Canadian Museum of Nature (collection ID#'s 80-505, 80-507, 81-1110).

MORPHOLOGICAL MEASUREMENTS

Morphological and meristic measurements were made on approximately 50 individuals per location. All fish were frozen to facilitate rapid collection and preservation. Within six months of collection 15 morphological and meristic traits were measured (see Figure 2), including fork length (tip of snout to caudal fork), eye diameter (diameter of orbit along body axis), upper jaw length (tip of jaw to end of maxilla), head depth (depth at opercular insert), head length (tip of head to posterior edge of operculum), pre-dorsal length (tip of upper jaw to anterior insert of dorsal fin), pre-pelvic length (tip of upper jaw to anterior insert of pelvic fin), pre-pectoral length (tip of upper jaw to anterior insert of pectoral fin), pelvic-pectoral fin distance (distance from pelvic insert to pectoral fin insert), adipose-fork distance (distance from adipose insert to fork of the caudal fin), pelvic-fork distance (distance from pelvic insert to fork of the caudal fin), dorsal-pectoral distance (distance from dorsal fin insert of the pectoral fin), (distance from dorsal fin insert of the pelvic fin), number of pelvic fin rays, number of pectoral fin rays, and number of dorsal fin rays. Gill raker measurements (i.e. counts and functional distances) were made on 2002-2003 specimens but not included due to concerns over consistency with previous studies. All linear measurements were made with Vernier callipers. Co-variation due to size was removed by linear regression against fork length and residuals were used in subsequent analysis. Principle Component Analysis and Quadratic Discriminant Function Analysis (Systat v11) was used to examine significant variation in form and the success at which the size based morphs could be identified.

MOLECULAR GENETIC ANALYSIS

DNA was extracted following the protocol of Elphinstone *et al.* (2003), modified to work with a 96-well filter plate and automated on a robotic liquid handling system (Perkin Elmer). Ten tetranucleotide microsatellite loci were used following Coulson et al. (2006). Individuals were genotyped using PCR conditions of 5- or 10-μL volumes containing 20-100 ng DNA, 1.5 mM $MgCl_2$, 80uM each dNTP, 0.5 U *Taq* DNA polymerase (New England Biolabs), 0.3 μM each primer (forward primers were end-labelled with Hex, or Rox dye), and 10x PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl). Two temperature profiles were used for touchdown in order to allow for the possibility of multiplex PCRs. Touchdown PCR conditions were as follows: 94°C for 2 min, followed by 4-5 cycles of 94°C for 30 s, program specific touchdown annealing temperatures minus 1°C per cycle for 30 s, 72°C for 30 s, followed by 25-26 cycles where the T_a was held constant at 4°C below the starting temperature. A final extension was held at 72°C for 5 min. Reactions were run on Eppendorf thermocyclers and imaged on an FMBioII system (Hitachi Genetic Systems). See Coulson et al. (2006) for further details. Data were checked for the presence of null alleles and scoring errors using MICRO-CHECKER (van Oosterhout et al., 2004). Genetic polymorphism was quantified by examination of the number of alleles, and observed and expected heterozygosity using GENETIX (version 4.05.2, Belkhir *et al.* 2004) Tests for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were done using FSTAT (version 2.9.3.3, Goudet, 1995). F-statistics and significance were calculated using MSA (Dieringer and Schlötterer 2003)

Bayesian clustering was performed using STRUCTURE ver.2.0 (Pritchard et al. 2000), and tested the presence of multiple discrete populations. This approach uses assumptions of HWE and linkage equilibrium among loci, introduces population structure, and assigns populations that are not in linkage disequilibrium using a MCMC algorithm to estimate the number of populations (K). The algorithm was run three times for each K to ensure convergence of values, and with a burn in of 250,000 reps, 500,000 reps after burn in, and K ranging from 1 to 6. In addition to K, STRUCTURE estimated a Q-value or admixture coefficient for each individual reflecting the portion of each multilocus genotype which was assigned to a given cluster or population. In instances where hybridization between two groups occurs, intermediate Q-values may represent hybrid individuals. Associations between Q-values and body size were explored using least squares linear regression.

The effective number of individuals contributing to each sample (i.e. N_e) was estimated using the linkage disequilibrium method (Hill 1981) for each of the samples using the program LDNE (Waples & Do 2008). LDNE estimates the number of effective of breeders contributing to a sample and applies a correction to account for sample sizes less than the actual effective size. Allele frequencies near one or zero can affect the value of N_e (Waples 2006) and as such three exclusion criteria were applied (0.01, 0.02, and 0.05). 95% confidence intervals for N_e were calculated using the jackknife approach implemented in LDNE.

RESULTS

MORPHOLOGICAL VARIATION

Size of spawning individuals were available for all samples from 1980-2010 (See Table 1). In years where only standard length or total length was available these were converted to fork length using a linear regression based on the association among all length measures (authors unpublished data). In all years there were obvious peaks in abundance at lengths greater and less than 170 mm (Figure 3). In addition, the 1991, 2002, and 2003 lower size class (<170 mm) seemed to include two independent size groups, one peaking at 140-150 and the other at 100-110 (Figure 3). Without accurate age data, it is impossible to compare age distributions among these groups. Furthermore, attempts to sample the large and intermediate size groups in 2010 were unsuccessful preventing comparison of these size classes with current samples. Gill raker counts were available from 1980-81, 90-91, and 2010 (Figure 4). Two modes were apparent in each sample period with the lower mode average ranging from 31.2-32.8 gill rakers and the upper mode average from 34.2-35.9 (Figure 4, 5). The values in 1990-1991 were higher for both the low and high gill raker classes. The transition from high to low gill raker counts seemed to coincide with approximately 170 mm fork length (Figure 5).

Multivariate analysis of morphological variation was examined using Principle Component Analysis (PCA) and Quadratic Discriminant Function Analysis (DFA). PCA of morphometric data, including size and other detrended variables for 2002-2003 revealed clear size based clusters (Figure 6a). PC1 was significantly correlated with fork length (R=-0.991, p-value <0.001) and easily separated the large size group from smaller individuals (<170 mm). Two clusters representing small and intermediate size individuals were also apparent (Figure 6a). PCA without size revealed a collapse of cluster structure and significant overlap among the size groups (Figure 6b). QDFA with size assigned individuals to two (<170 and >170 mm) and three (<120, 120-170, >170 mm) size groups with 99% and 98% accuracy respectively (Table 2.). In both cases the removal of size resulted in decline in accurate assignment to 69 and 61% respectively (Table 2.).

GENETIC ANALYSIS

Average microsatellite heterozygosities were lower in large morph (0.73) than in small morph (0.76) though this difference was not significant (ANOVA, p=0.266). Average number of alleles did significantly differ (ANOVA, p=0.009) between the large (N_a=8.9) and small (N_a=10.9) morphs. Pairwise F_{ST} values ranged from 0 to 0.067, averaged 0.026 across all sample comparisons, and the global estimate was 0.025 (Table 3). Approximately 85% of all significant comparisons (n=19) were between the large and other forms. Pairwise F_{ST} comparisons by size class again revealed only significant comparisons involving the large form with the exception of comparisons with the 2002 intermediate size sample which differed from the 2003 and 2010 small samples (Table 4).

Genetic associations among individuals were examined using Principle Coordinate Analysis and Bayesian clustering using STRUCTURE. Principle Coordinate Analysis revealed some independence among the locations and size groups but there was significant overlap (Figures 7-9). In particular, in 2002 significant overlap was observed among the locations but this appeared driven by small individuals possessing the ‘large” genotype (i.e., high PC1 values, Figure 8). When all samples were analyzed together, the 2002 samples were highlighted as divergent on PC1 (Figure 10a). The subsequent removal of samples from 2002 restored the association between PC1 and body size (Figure 10b).

Clustering using STRUCTURE supported the presence of two discrete groups in each comparison with the ΔK values peaking at two clusters (Figure 11, 12). In 1990, 2003, and 2010 the clusters observed represented the large and small size classes (Figure 12). In 2002, a number of large individuals in Trout Lake Stream and Mill Lake Stream were predominantly hybrids or the small genetic form (Figure 12b). STRUCTURE analysis with all samples indicated consistent clustering of brook and stream samples from all time periods, again with the exception of the 2002 stream samples. The range of Q-values observed varied significantly with body size (Figure 13). Overall there was a sharp decline in Q-value at ~170 mm associated with the transition to the large size morph (Figure 13 b,c). At smaller sizes (<170 mm) a range of Q-values were present indicating that some small individuals in all samples were genetically similar to the large morph. The frequency distribution of Q-values for the large and small forms indicated a greater portion of small individuals with intermediate Q-values than present in the large morph (Figure 14.). STRUCTURE analysis with fish <170 mm, confirmed the presence of genetically large individuals, yet Q-values displayed no association with size in this length range (Figure 15) suggesting these individuals were phenotypically the small morph.

N_e ESTIMATION

Estimation of N_e revealed values ranging from 33.5 to 354.0 (Table 5) and there was no association with morph. However given the inability to estimate the upper confidence intervals using the jackknife procedure these values should be interpreted cautiously. The consistency of the estimates and the tendency for all to lie in the 10s-100’s of individuals might support the hypothesis that the actually values are within this range but again caution is recommended and further analysis is required to refine these estimates.

DISCUSSION

We examined the presence of phenotypic and genetic divergence within Lake Utopia smelt and observed significant variation congruent with earlier work hypothesising two distinct forms (e.g., Lanteigne and McAllister 1983; Taylor and Bentzen 1993a). These two forms display moderate genetic isolation at multiple microsatellite loci and clear morphological variation associated with previously documented alternate trophic strategies (Kendall 1927; Bajkov 1936; Delisle 1969). Within the microphagous form, biomodality based on size was observed however its ecological or evolutionary significance will require further analysis. Consistent evidence of hybridization, visible in the presence of genetic hybrids and phenotypically small individuals with “large morph” genotypes supports a hypothesis of continued gene flow. The persistence and stability of this pair likely depend on the selective landscape as changing biotic and abiotic conditions could destabilize the complex (e.g., Taylor et al. 2006). Further study of the rate of hybridization, the fate of hybrids, and the selective landscape associated within the lake are needed to evaluate the status, stability, and allow continued conservation of smelt diversity in Lake Utopia.

The presence of intraspecific diversification associated with alternate feeding specializations has been well documented across a range of taxa (e.g., Nosil et al. 2005; Barrett et al. 2009). In Lake Utopia smelt, morphological diversity has an extensive history of study. Lanteigne and McAllister (1983) used phenotypic variation to show little statistical overlap between the large and small forms and based their key for morph identification of the forms on length, gill raker counts, and eye diameter. Our observation that phenotypic divergence among the large and small forms was primarily associated with size at maturity and gill raker counts is consistent with these earlier studies. The discriminatory power reported here for identifying morphs using phenotypic data is similar to studies of other sympatric smelt pairs. Delisle (1969) examined differences in morphometric and meristic traits among a sympatric large and small morph smelt pair in Lake Heney, and achieved 100% accuracy in correct assignment using a discriminant function. Similarly, Copeman (1977) used 32 mensural and meristic traits and reported 100% success in separating the small form from the large form three separate locations. Accordingly the degree of phenotypic variation observed in Lake Utopia smelt would appear on par with previous descriptions of the recorded seven sympatric smelt pairs described to date.

From a management viewpoint, a critical issue is the identification of the discrete stocks present within the lake. Elsewhere, studies (Kendall 1927; Delisle 1969; Copeman 1977) have consistently supported the conclusion of two divergent forms, a large macrophagous form and a small microphagous form. As both our genetic and morphological data (linear measures and gill raker counts) support the presence of two forms in Lake Utopia, this would seem the most parsimonious description of the stock complex. Moreover, many of the within morph genetic comparisons were non-significant again supporting the presence and stability of morph based isolation. However, Curry et al. (2004) report morphological and genetic evidence for three discrete forms within Lake Utopia based on extensive sampling of spawning runs in 1998-2000. Their third intermediate form may coincide with the phenotypic subdivision we observed in our small morph, but this remains to be tested.

In comparison to our genetic results, Curry et al. (2004) present a contrasting picture of the genetic structure of Lake Utopia smelt. As observed here, Curry et al. (2004) report the large or giant form was the most divergent. However the magnitude of divergence observed differs among the studies as they report an average overall F_{ST} value of 0.091 contrasting an average divergence here of F_{ST} = 0.026. Interestingly, Curry et al. (2004) report relatively high F_{ST} values among spawning sites within a form (e.g., small form averaging 0.05) whereas here we report largely non-significant divergence within a form among sites. Furthermore the topology of the dendrogram reported by Curry et al. (2004) displays no association with form and contrasts our clustering analyses which consistently identify morph and size as the dominant influences on genetic structure. The STRUCTURE analysis presented here, which resolves the most likely number of populations through the clustering of individuals, provides perhaps the most compelling support for two discrete forms clearly associating small and large individuals across all samples. This stability in population divergence across two decades indicated from the clustering analysis strongly supports the hypothesis that the primary division within Lake Utopia smelt is into two size-based forms. Further work is needed to address discrepancies between Curry et al. (2004) and this work and issues associated with unresolved diversity and stability with the small form. Also of note in our study is the distinction of the 2002 samples from all other samples as observed on the clustering plots as well as the increased rate of misassignment in Trout and Mill Lake Stream individuals in 2002. This is contrary to the general trend of more frequent intermediate Q-values in the small form than the large. These samples may indeed represent hybrids, yet the exact cause and nature of this occurrence will require further evaluation.

Age data will be required to fully interpret the status of the small and intermediate size morphs. Recent work on growth rates of the various morphs of smelt in Lake Utopia indicates the forms start diverging almost immediately following hatch (Shaw and Curry 2005). Comparisons of the average age of spawning individuals both within Lake Utopia and among other sympatric pairs tend to show significant variability. Preliminary aging on the small form sampled here from 2010 indicates most individuals are 3-4 years old (R. Bradford unpublished data). Curry et al (2004) report an average age of maturity of 2.8 years for both large and small morph spawning groups in Lake Utopia. In Lake Heney, (Delisle 1969) reported that the small morph grew slowly and reached maturity at 2 years of age whereas the large form grew quickly and most matured at 3 years of age. Similarly, Copeman (1977) reported that spawning individuals from Heney and Green lakes consisted mainly of 2 year olds for the small morph and 3-5 year olds for the large morph. Together the available estimates suggest large regional variation in growth and age of spawning individuals though clear differences in growth, age at maturity, and longevity seem likely.

Ultimately, the stability and persistence of sympatric pairs depends on the continued balance of hybridization and the selective removal of hybrid individuals. The possibility that sympatric forms may collapse to a single population has received significant attention in the literature (Frost 1965; Coyne and Orr 2004; Taylor et al. 2006) and may be most often associated with environmental change and young post glacial lineages (e.g., Taylor et al. 2006). For example, Enos Lake threespine stickleback (*Gasterosteus aculeatus*) was a morphological biomodal pair of benthic and limnetic populations which collapsed possibly following the introduction of an crayfish species in the 1990's (Taylor et al 2006). In general, the stability of a sympatric pair may be compromised if the relative fitness of hybrids increases or environmental cues associated with the maintence of divergence change. In the Lake Utopia pair, as the observed genetic differentiation appears stable over two decades and moderate genetic differentiation is present, selection appears efficient at removal of hybrids. Moreover the associations between genetic structure and body size indicate stability associated with size in the morphs present. As we demonstrate hybrids are likely present, the future maintence of the pair will depend on the relative fitness of hybrid individuals. Undoubtedly, a better understanding of the role of selection and hybridization dynamics will be required if management and conservation efforts are to effectively manage for the stability of the Lake Utopia sympatric smelt pair.

SUMMARY

We present genetic and phenotypic evidence for the presence of two forms of rainbow smelt in Lake Utopia New Brunswick which can be visible in samples spanning the last two decades. These forms are identifiable genetically, allowing management efforts to base census estimates and further data collection on a genetic basis. We present evidence of hybridization among the forms and suggest the continued stability of this pair will depend on environmental stability and better understanding of the selective maintenance of diversity within the lake. Such knowledge will be central to the management and conservation of this fragile sympatric pair.

REFERENCES

Barrette, M., Daigle, G., and Dodson, J. 2009. Intraspecific vicariant history and the evolution of adaptive morphological diversity within species. Biological Journal of the Linnaean Society 97: 140-151.

Bajkov, A.D. 1936. Western Atlantic smelt, *Osmerus eperlanus mordax* (mitchill) and its landlocked forms. In: A.D. Bajkov (Ed.) Investigations on Smelt in Chamcook Lake, N.B. Canada. p. 6.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. 2004. Genetix 4.02, logiciel sous WIndows pour la genetique des populations. Laboratoire Genome, Populations, Interactions. CNRS UMR 5000. Universite de Montpellier II, Montpellier, Fr. pp. -.

Bell, M. 1976. Reproductive character displacement in threespine sticklebacks. Evolution 30: 847-848

Bell, M. 1989. Morphogenesis and Evolution. Quarterly Review of Biology 64: 345-346

Bernatchez, L., Chouinard, A., and Lu, G. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. Biological Journal of the Linnean Society 68(1): 173-194.

Claytor, R.R., and Verspoor, E. 1991. Discordant phenotypic variation in sympatric resident and anadromous Atlantic salmon (*Salmo salar*) populations. Canadian Journal of Zoology 69: 2846-2852.

Copeman, D.G. 1977. Population differences in rainbow smelt, *Osmerus mordax*: Multivariate analysis of mensural and meristic data. Journal of Fisheries Research Board of Canada, Volume 34: 1220-1229pp.

Copeman, D.G., and McAllister, D.E. 1978. Analysis of the effect of transplantation on morphometric and meristic characters in lake populations of the rainbow smelt, *Osmerus mordax* (Mitchill). Environmental Biology of Fisheries, 3(3), 253-259, (1978): 3.

Coulson, M.W., Paterson, I.G., Green, A., Kepkay, R., and Bentzen, P. 2006. Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). Molecular Ecology Notes 6(3): 942.

Coyne, J., and Orr, H.A. 2004. Speciation. Sinauer Associates, Inc. Sunderland, Massachusetts.

Curry, R.A., Currie, S.L., Bernatchez, L., and Saint-Laurent, R. 2004. The rainbow smelt, Osmerus mordax, complex of Lake Utopia: threatened or misunderstood? Environ. Biol. Fishes 69(1-4): 153-166.

Delisle, C. 1969. Écologie, croissance et comportement de l'éperlan du lac Heney, comté de Gatineau ainsi que la répartition en eau douce au Québec, Department of Biology, University of Ottawa, Ottawa.

Dieringer, D., and Schlötterer, C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3(1): 167-169.

Elphinstone, M.S., Hinten, G.N., Anderson, M.J., and Nock, C.J. 2003. An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. Molecular Ecology Notes 3(2): 317-320.

Frost, W.E. 1965. Breeding habits of WIndermere charr, Salvelinus willughbii (Gunther), and their bearing on speciation of these fishes Proc R Soc Lond B Bio 163: 232-284.

Goudet, J. 1995. Fstat Version 1.2: A computer program to calculate F-statistics. Journal of Heredity 86: 485.

Green, C.W. 1930. The smelts of Lake Champlain; pp. 105-129. In: A Biological Survey of the Champlain Watershed. A supplement to the 19th Annual Report of the New York Conservation Department for 1929.

Hendry, A., Huber, S., De León, L., Herrel, A., and Podos, J. 2009. Disruptive selection in a bimodal population of Darwin's finches. Proceedings of the Royal Society B. Biological Sciences 276: 753-759.

Hendry, A., Grant, P., Grant, B., Ford, H., Brewer, M., and Podos, J. 2006. Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. Proceedings of the Royal Society B. Biological Sciences 273: 1887-1894.

Hill, W.G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetical Research 38: 209-216.

Kendall, W.C. 1927. The smelts. Fishery Bulletin of the Fish and Wildlife Service 42: 217-325.

Lanteigne, J., and McAllister, D. 1983 The pygmy smelt, *Osmerus* spectrum Cope, 1870, A forgotten sibling species of Eastern North American Fish., Ichthyology Section, National Museum of Natural Sciences, Canada., Ottawa.

Lu, G., and Bernatchez, L. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. 53: 1491.

McPhail, J.D. 1994. Speciation and the evolution of reproductive isolation in the sticklebacks(Gasterosteus) of southwestern British Columbia; pp. In: M.A. Bell and S.A. Foster (eds.) The Evolutionary Biology of the Threespined Stickleback. Oxford University Press, Oxford, UK

Nellbring, S. 1989. The ecology of smelts (genus *Osmerus*): A literature review. Nordic Journal of Freshwater Research 65: 116-145.

Nosil, P., Vines, T.H., and Funk, D.J. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59(4): 705-719.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000/6. Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Rogers, S., and Bernatchez, L. 2006 The genetic basis of intrinsic and extrinsic postzygotic isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*) Journal of Evolutionary Biology 19: 1979 - 1994

Rupp, R.S. 1959. Variation in the life history of the American smelt in inland waters of Maine. Trans Am Fish Soc 88(241-252).

Rupp, R.S., and Redmond, M.A. 1966. Transfer studies of ecologic and genetic variation in American smelt Ecology 47: 253–259.

Schluter, D., and McPhail, J. 1992. Ecological character displacement and speciation in sticklebacks. American Naturalist 140: 85–108.

Shaw, J., and A. Curry. 2005. Lake Utopia Rainbow Smelt Report 2004. Prepared for the New Brunswick Wildlife Trust Fund. New Brunswick Cooperative Fish and Wildlife Research Unit. Report 05-05.

Taylor, E., and Bentzen, P. 1993. Molecular genetic evidence for reproductive isolation between sympatric populations of smelt *Osmerus* in Lake Utopia, south-western New Brunswick, Canada. Molecular Ecology 2: 345.

Taylor, E., Boughman, J., Groenenboom, M., Schluter, D., Sniatynski, M., and Gow, J. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair Molecular Ecology 15 343-355.

Taylor, E.B., and Bentzen, P. 1993a. Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47(3): 813-838.

Taylor, E.B., and Bentzen, P. 1993b. Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47(3): 813-838.

Waples, R.S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci Conservation Genetics 7(2): 167-184.

Waples, R.S., and Do, C. 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. Mol. Ecol. Resources 8: 753-756.

Table 1. Sample collection locations and characteristics of available data for smelt from Lake Utopia, New Brunswick. See Figure 1 for sample locations. For samples where some morphological data was available but insufficient for multivariate analysis (e.g., length), this has been noted by "limited".

Sample ID	Year	Location / Sample details	Sample size	Genetic data	Gill raker counts	Morphometric data	Age information
1	1980/1981	Large and small morph (Lanteigne and McAllister 1983)	71	No	Yes	limited	Yes
2	1990/1991	Large and small morph (Taylor and Bentzen 1993a)	67	Yes	Yes	Limited	No
4	2002	Mill Brook	84	Yes	No	Yes	No
5	2002	Unnamed Brook	65	Yes	No	Yes	No
6	2002	Scout Brook	45	Yes	No	Yes	No
7	2002	Trout Lake Stream	20	Yes	No	Yes	No
8	2002	Mill Lake Stream	54	Yes	No	Yes	No
9	2003	Unnamed Brook	91	Yes	No	Yes	No
10	2003	Scout Brook	96	Yes	No	Yes	No
11	2003	Mill Lake Stream	45	Yes	No	Yes	No
12	2003	Trout Lake Stream	45	Yes	No	Yes	No
13	2010	Unnamed Brook	48	Yes	Yes	Limited	Yes
14	2010	Mill Brook	49	Yes	Yes	limited	yes

Table 2. Correct assignment from Quadratic Discriminant Function Analysis of morphometric data from Lake Utopia rainbow smelt sampled in 2002-2003. Size groupings are based on dominant modes, and analysis is preformed with detrended variables (i.e. residuals of regression with length) with or without fork length. See Figure 2 for traits examined and Methods for further details.

		Size classes			
# of groups	**Size included**	**Small** (<120 mm)	**Intermediate** (>120 and <170 mm)	**Large** (>170 mm)	**Total**
Two	Y	100		96	**99%**
	N	70		65	69%
Three	Y	99	97	97	**98%**
	N	58	69	54	61%

Table 3. Pairwise F_{ST} values (above diagonal) and p-values (below) for all samples of rainbow smelt from Lake Utopia (1990-2010). Significance calculated using permutation. See Table 1 and Figure 1 for sample locations and characteristics.

	Large 1990	Small 1990	MillBk 2002	UnBk 2002	ScoutBk 2002	TroutLkStr 2002	MillLkStr 2002	UnBk 2003	ScoutBk 2003	MillLK 2003	TroutLkStr 2003	Unbk 2010	Smbk 2010
1.Large 1990		**0.044**	**0.044**	**0.050**	**0.058**	**0.039**	**0.047**	**0.057**	**0.062**	**0.029**	**0.025**	**0.061**	**0.061**
2. Small 1990	<0.001		0.002	0.002	**0.024**	**0.026**	**0.055**	**0.021**	**0.022**	**0.062**	**0.066**	**0.017**	**0.024**
3. MillBk02	<0.001	n.s.		-0.001	0.005	-0.001	**0.034**	-0.002	0.003	**0.047**	**0.051**	0.003	0.000
4. UnBk02	<0.001	n.s.	n.s.		0.006	0.010	**0.038**	0.000	-0.001	**0.048**	**0.057**	0.009	0.001
5. ScoutBk02	<0.001	<0.001	n.s.	n.s.		-0.006	**0.022**	0.001	0.010	**0.054**	**0.048**	0.011	0.009
6. TroutLkStr02	<0.001	<0.001	n.s.	n.s.	n.s.		0.009	0.002	0.006	**0.041**	**0.037**	0.005	0.010
7. MillLkStr02	<0.001	<0.001	<0.001	<0.001	<0.001	n.s.		**0.030**	**0.036**	**0.034**	**0.031**	0.037	0.032
8. UnBk03	<0.001	<0.001	n.s.	n.s.	n.s.	n.s.	0.001		0.001	**0.042**	**0.042**	0.006	0.001
9. ScoutBk03	<0.001	<0.001	n.s.	n.s.	n.s.	n.s.	0.001	n.s.		**0.042**	**0.050**	0.006	0.001
10. MillLK03	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001		0.003	0.045	0.040
11. TroutLkStr03	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	n.s.		0.051	0.048
12. Unbk10	<0.001	<0.001	n.s.	n.s.	<0.001	n.s.	<0.001	n.s.	n.s.	<0.001	<0.001		0.002
13. smbk10	<0.001	<0.001	n.s.	n.s.	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001	<0.001	n.s.	

Table 4. Pairwise F_{ST} values (above diagonal) and p-values (below) for all size groups of rainbow smelt from Lake Utopia. Significance calculated using permutation. See Table 1 and Figure 1 for sample locations and characteristics.

	Small 2002	Intermediate 2002	Small 2003	Intermediate 2003	Small 2010	Intermediate 2010	Large 2002	Large 2003
Small 2002		-0.004	-0.002	-0.003	-0.002	-0.002	**0.037**	**0.041**
Intermediate 2002	n.s.		0.007	0.004	0.010	0.0131	**0.040**	**0.053**
Small 2003	n.s.	<0.001		0.003	0.002	0.006	**0.046**	**0.044**
Intermediate 2003	n.s.	n.s.	n.s.		0.005	0.0104	**0.030**	**0.033**
Small 2010	n.s.	<0.001	n.s.	n.s.		0.003	**0.037**	**0.039**
Intermediate 2010	n.s.	n.s.	n.s.	n.s.	n.s.		**0.058**	**0.058**
Large 2002	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001		**0.013**
Large 2003	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	

Table 5. Estimates of effective population size for rainbow smelt samples from Lake Utopia, New Brunswick from 2003 to 2010 by A) population (as calculated from STRUCTURE analysis) and B) sample. Estimates were calculated using the linkage disequilibrium method using LDNE (Waples and Do 2009). Confidence intervals were calculated using a jackknifing procedure, and infinity as the upper confidence interval for most estimates indicates a lack of power in estimation.

		Lowest allele frequency used		
Sample	**year**	**0.05**	**0.02**	**0.01**
A) By population				
Small Morph	2003	98.8 (17.1-∞)	137.7 (25.6-∞)	268.7 (40.8-∞)
	2010	59.9 (12.3-∞)	85.7 (22.5-∞)	90.5 (22.5-∞)
Large Morph	2003	62.0 (23.7-1364)	136.8 (43.6-∞)	141.6 (49.8-∞)
B) By sample				
Unnamed Brook	2003	41.6 (4.0-∞)	66.8 (14.0-∞)	87.5 (21.5-∞)
Scout Brook	2003	226.7 (28.6-∞)	343.5 (42.8 - ∞)	314.8 (73.4-∞)
Mill Lake Stream	2003	33.5 (13.1-278.8)	48.0 (20.0-587.6)	62.4 (25.6-∞)
Trout Lake Stream	2003	345.5 (154-∞)	354 (137.5-∞)	347 (145.6-∞)
Unnamed Brook	2010	130.4 (17.2-∞)	194.1 (30.4-∞)	223.7 (38.6-∞)
Smelt Brook	2010	54.2 (14.2-∞)	62.1 (16.7 - ∞)	104.1 (29.2-∞)

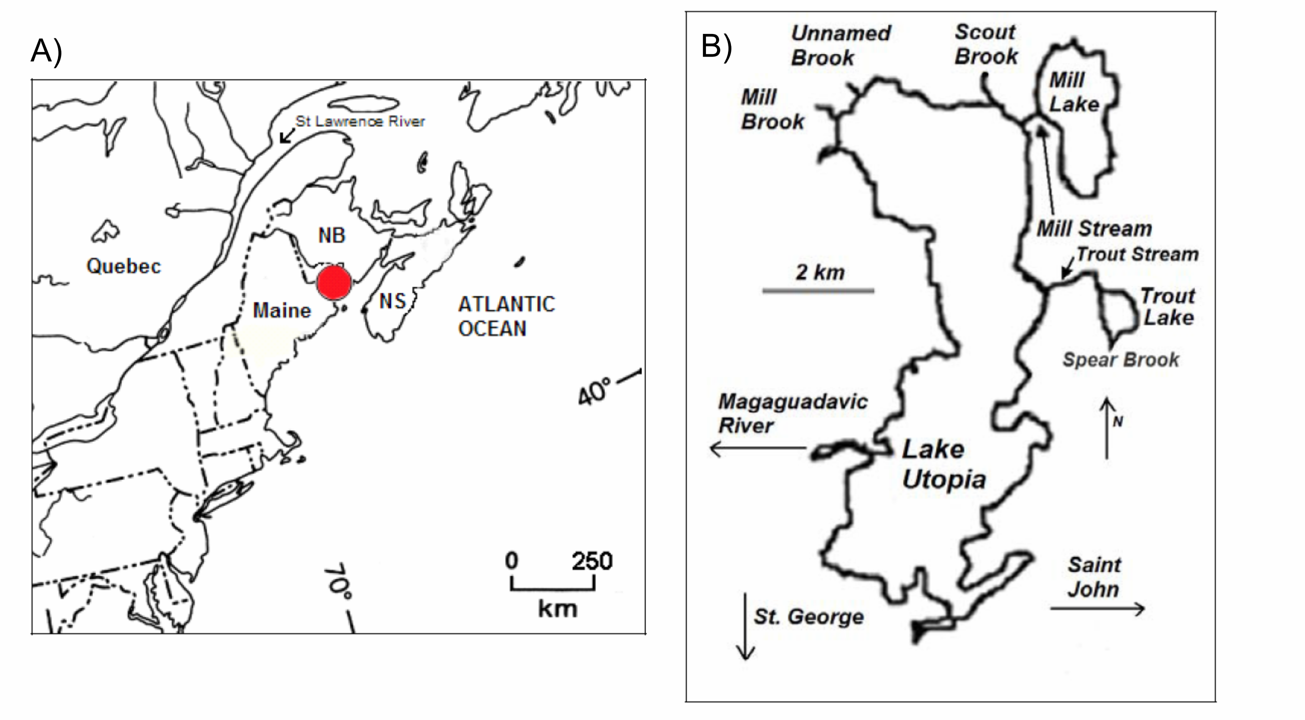


Figure 1. Map of (A) Lake Utopia with respect to eastern North America and (B) various sampled locations within the lake.

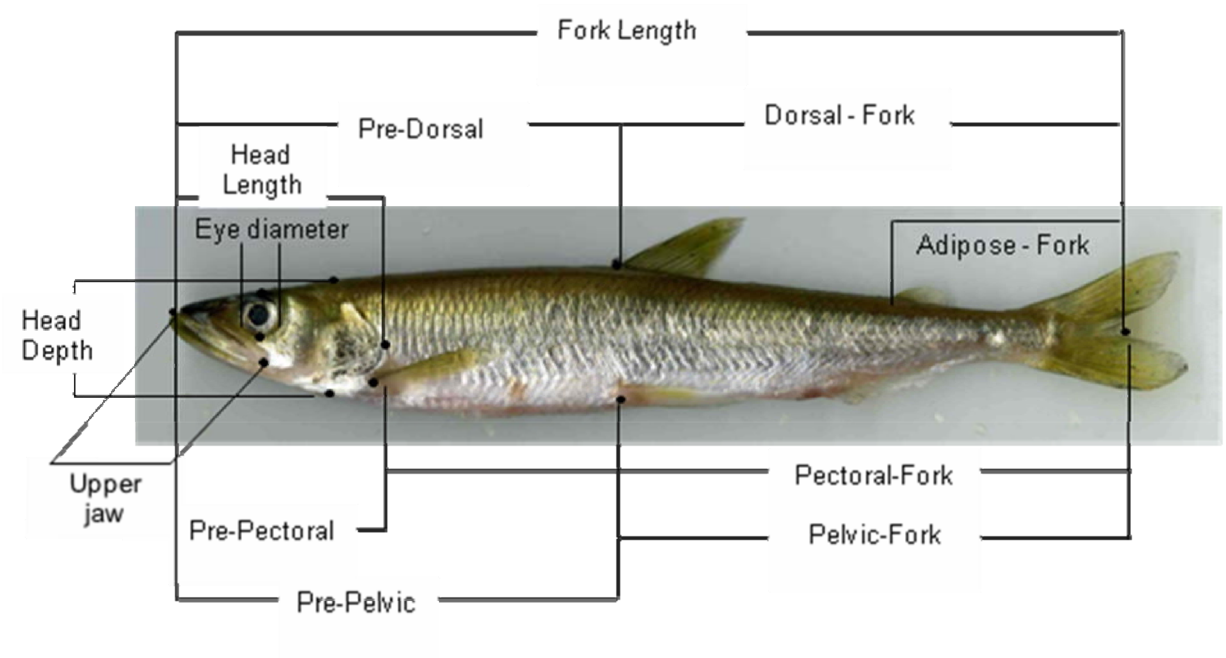


Figure 2. Linear morphometric measures used in analysis. See Methods for details and explanation.

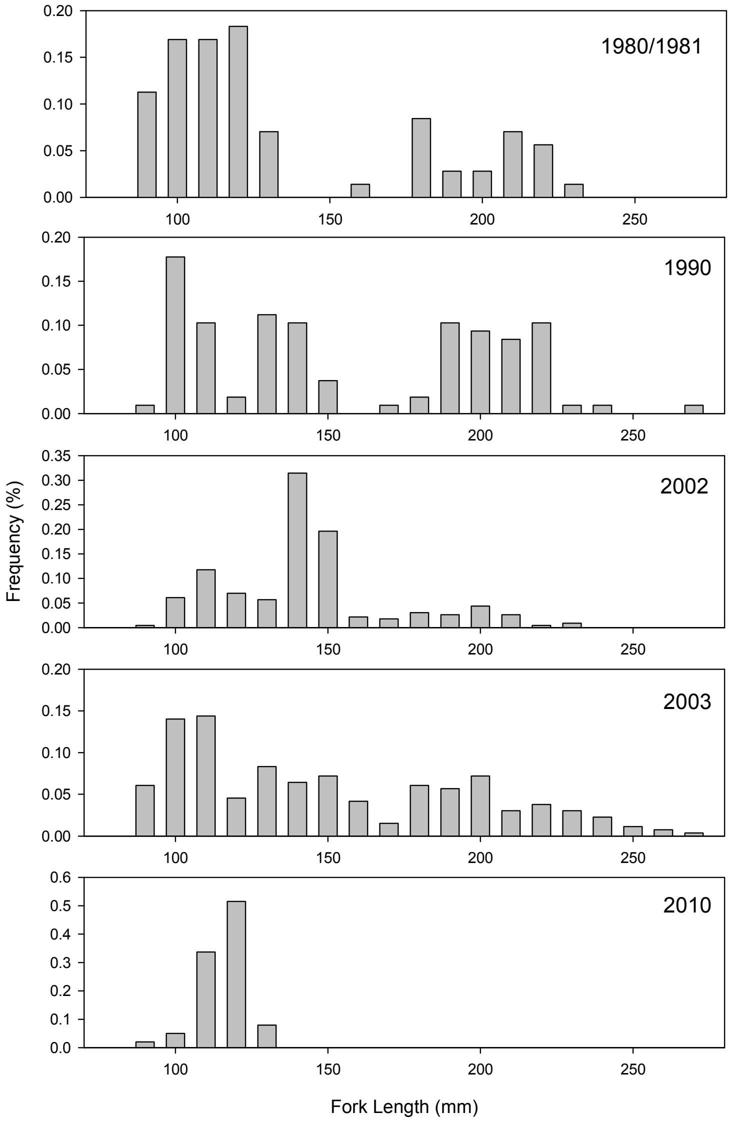


Figure 3. Frequency distribution of length data for samples from 1980-2010. See Table 1 for data sources. Note that no large morph individuals (i.e. streams) were sampled in 2010.

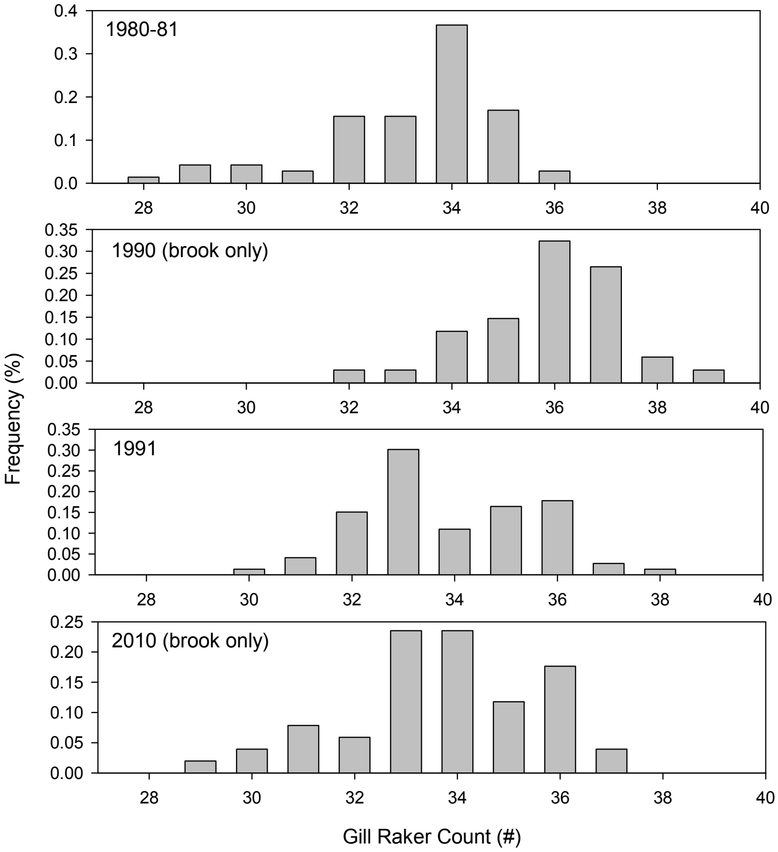


Figure 4. Frequency distribution of gill raker counts for samples from 1980-1981, 1990-1991, and 2010. See Table 1. for data sources. Note that no large morph individuals (i.e. streams) were sampled in 2010.

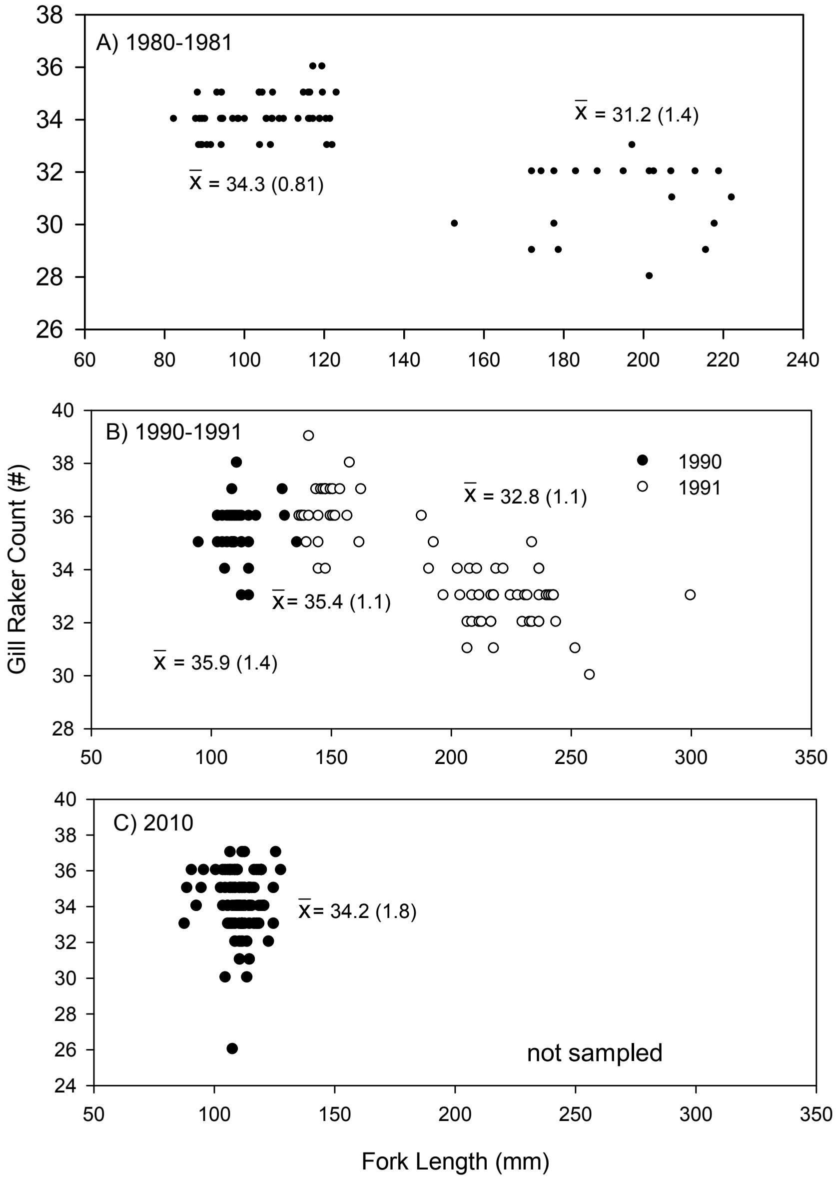


Figure 5. Gill raker count (#) and fork length (mm) for the samples of rainbow smelt from Lake Utopia from 1980-2010. Average (STD) values are given for large, intermediate (when present) and small morphs.

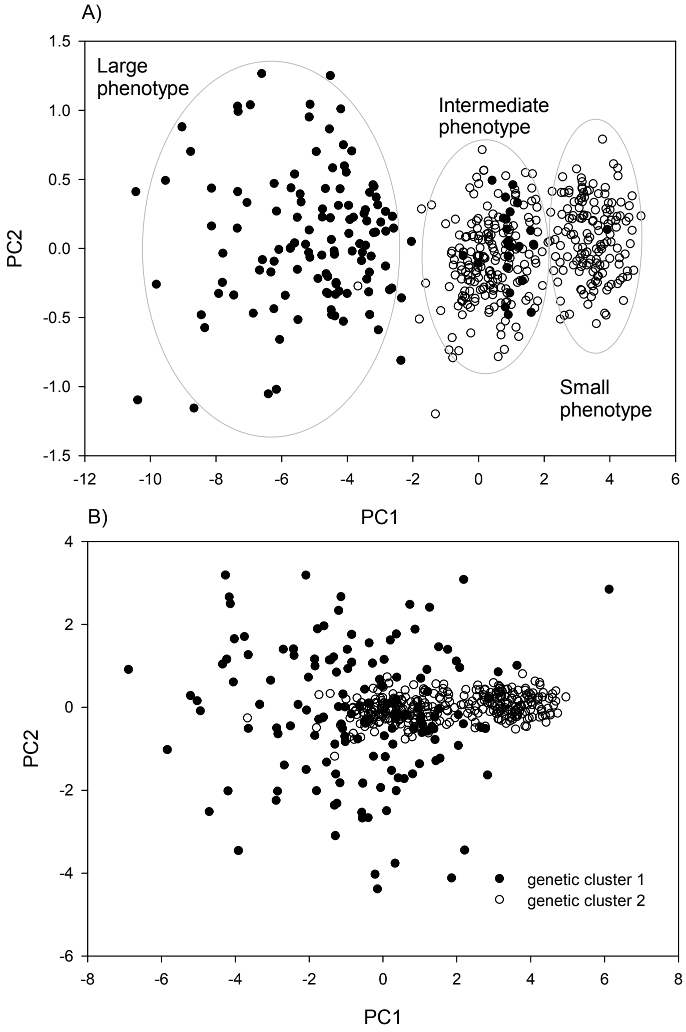


Figure 6. Principle component analysis of morphometric data with (A) and without (B) the effect of size. See Methods for analysis details. Groups in (A) based on average size class divisions from Figure 3.

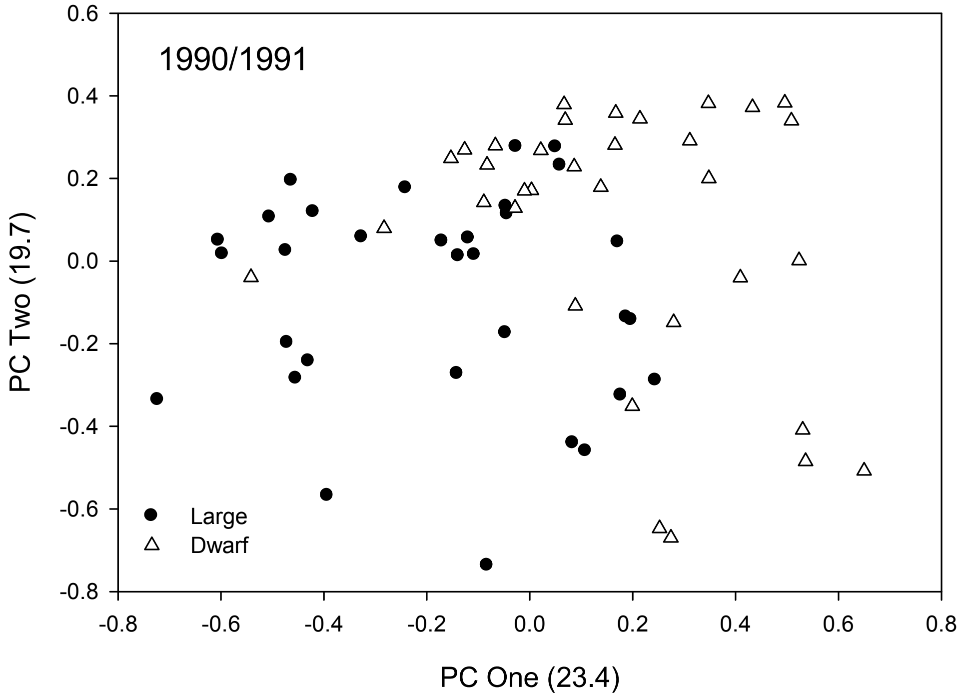


Figure 7. Principle Coordinates Analysis based on 10 microsatellite loci for large and small smelt sampled in 1990-1991. See Table 1 and Taylor and Bentzen (1993a) for sample details.

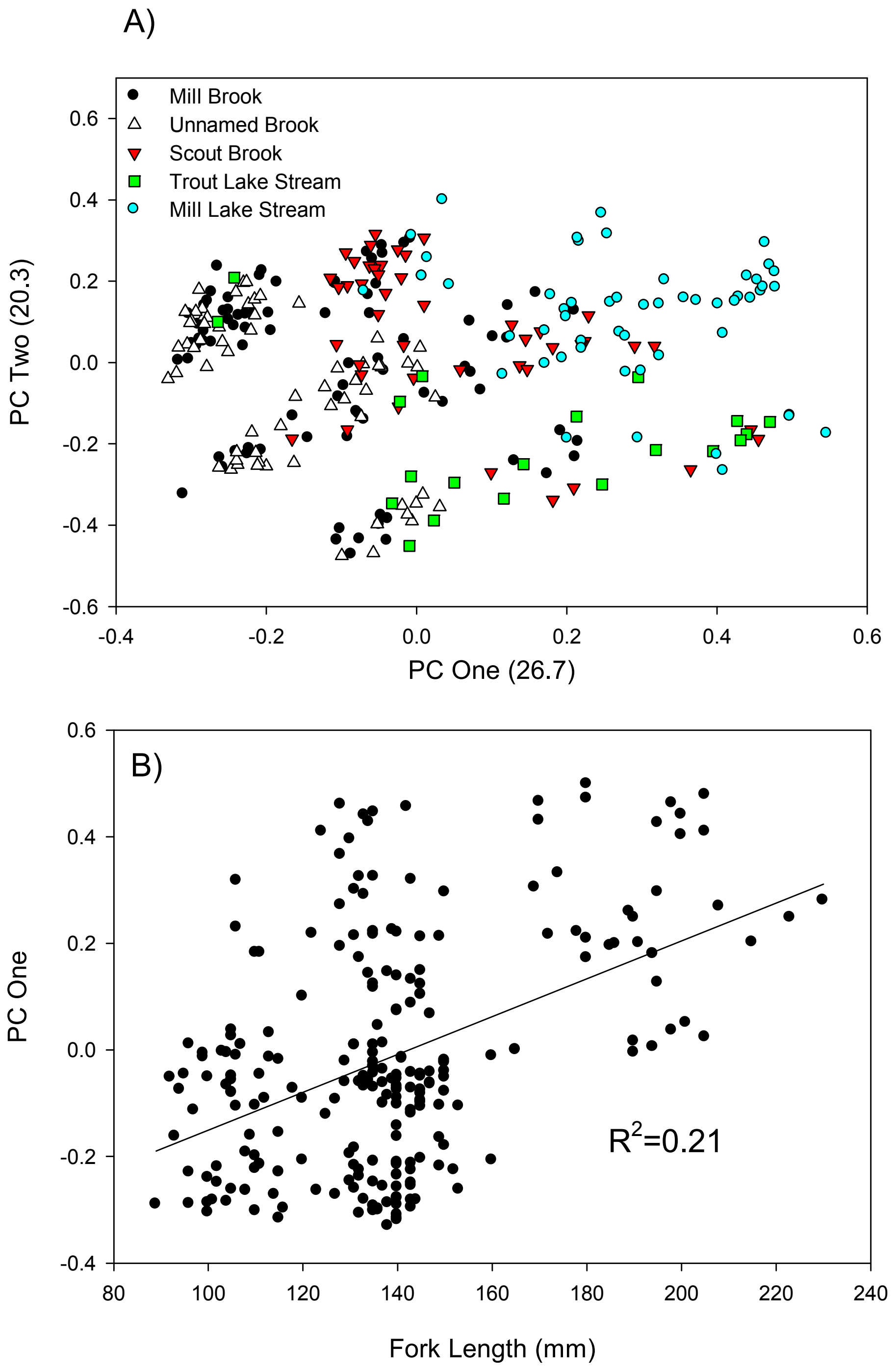


Figure 8. (A) Principle Coordinates Analysis based on 10 microsatellite loci for smelt sampled from five locations in Lake Utopia in 2002 .(B) Relationship between the first principle coordinate and fork length. See Table 1 and Figure 1 for sample details.

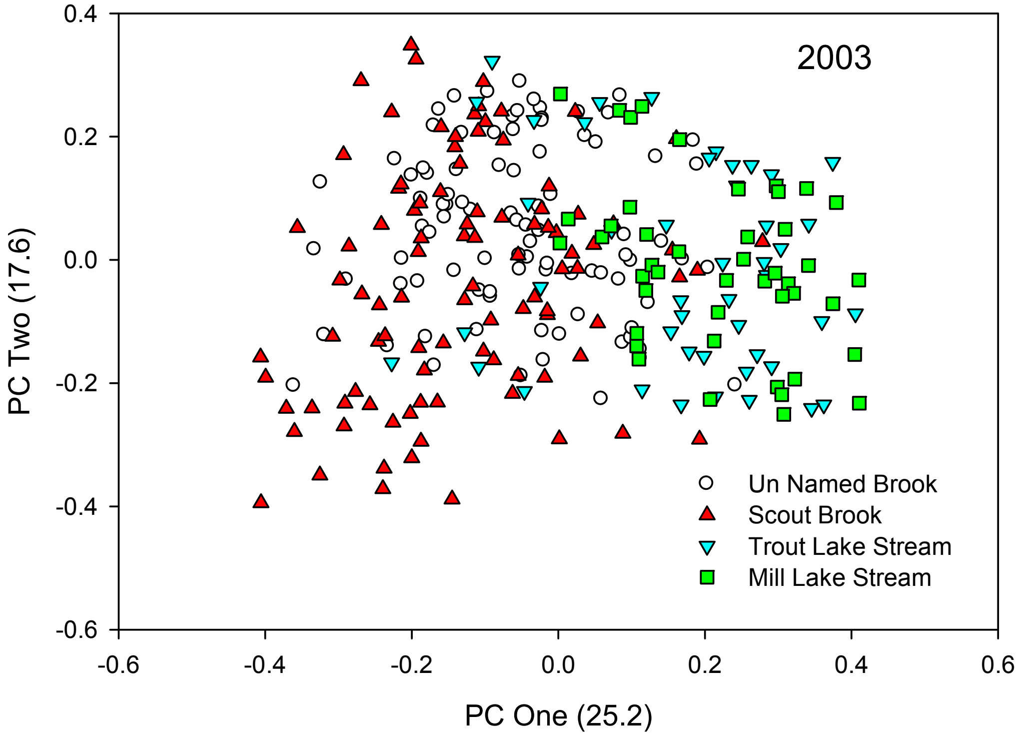
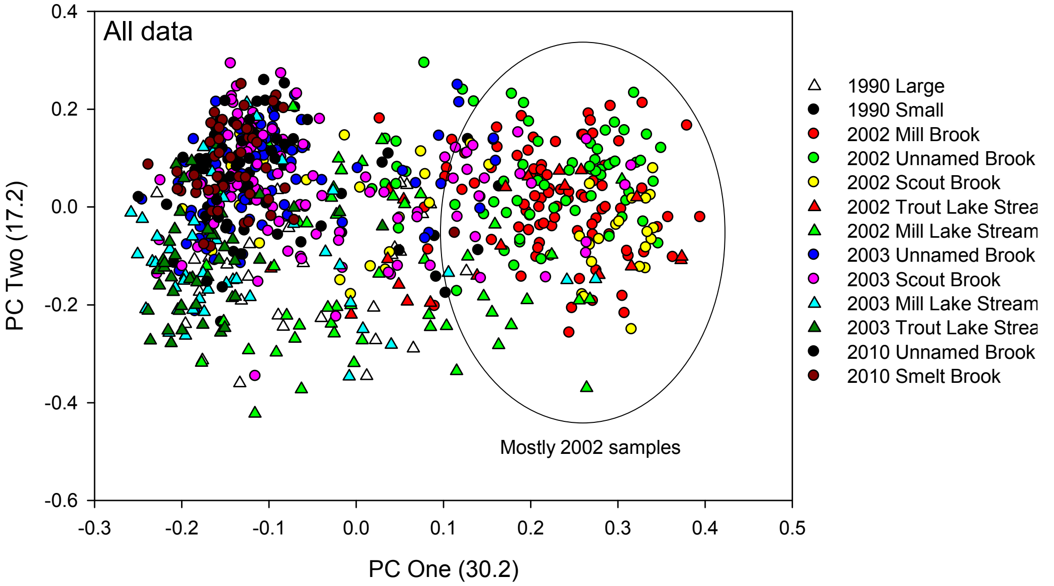


Figure 9. (A) Principle Coordinates Analysis based on 10 microsatellite loci for smelt sampled from four locations in Lake Utopia in 2003. See Table 1 and Figure 1 for sample details.



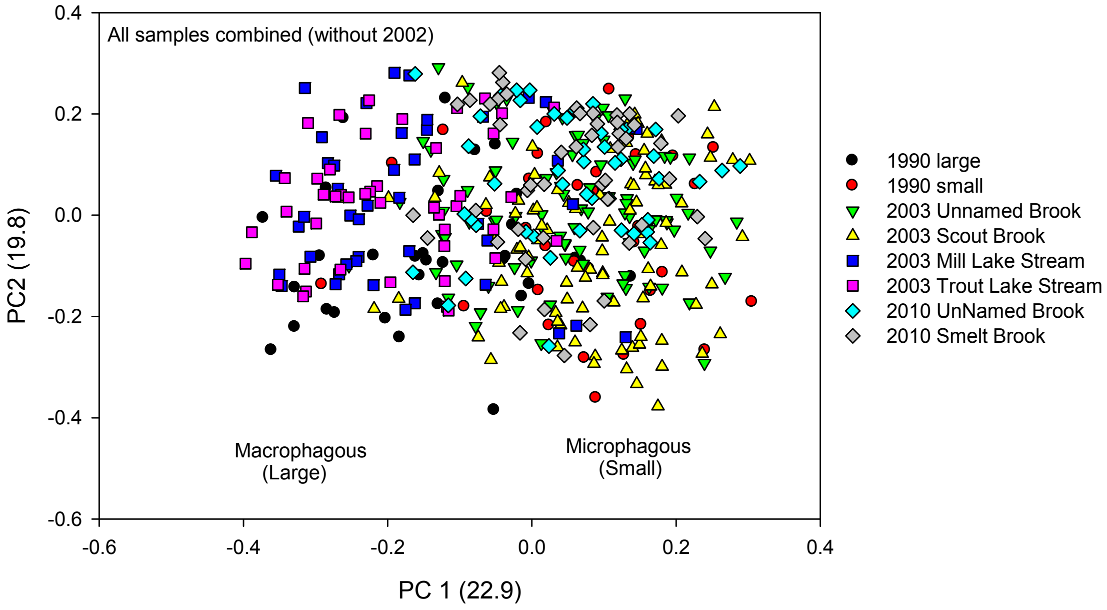


Figure 10. (A) Principle Coordinates Analysis of 10 microsatellite loci for smelt sampled from 1990-2010 in Lake Utopia (B) Principle Coordinates Analysis of smelt sampled from 1990-2010 in Lake Utopia excluding 2002 samples. See Table 1 and Figure 1 for sample details.

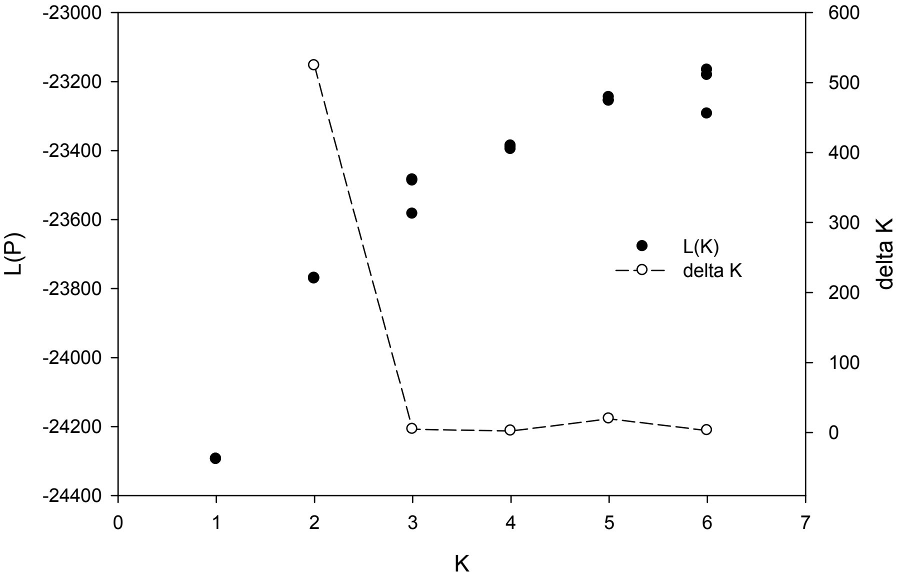
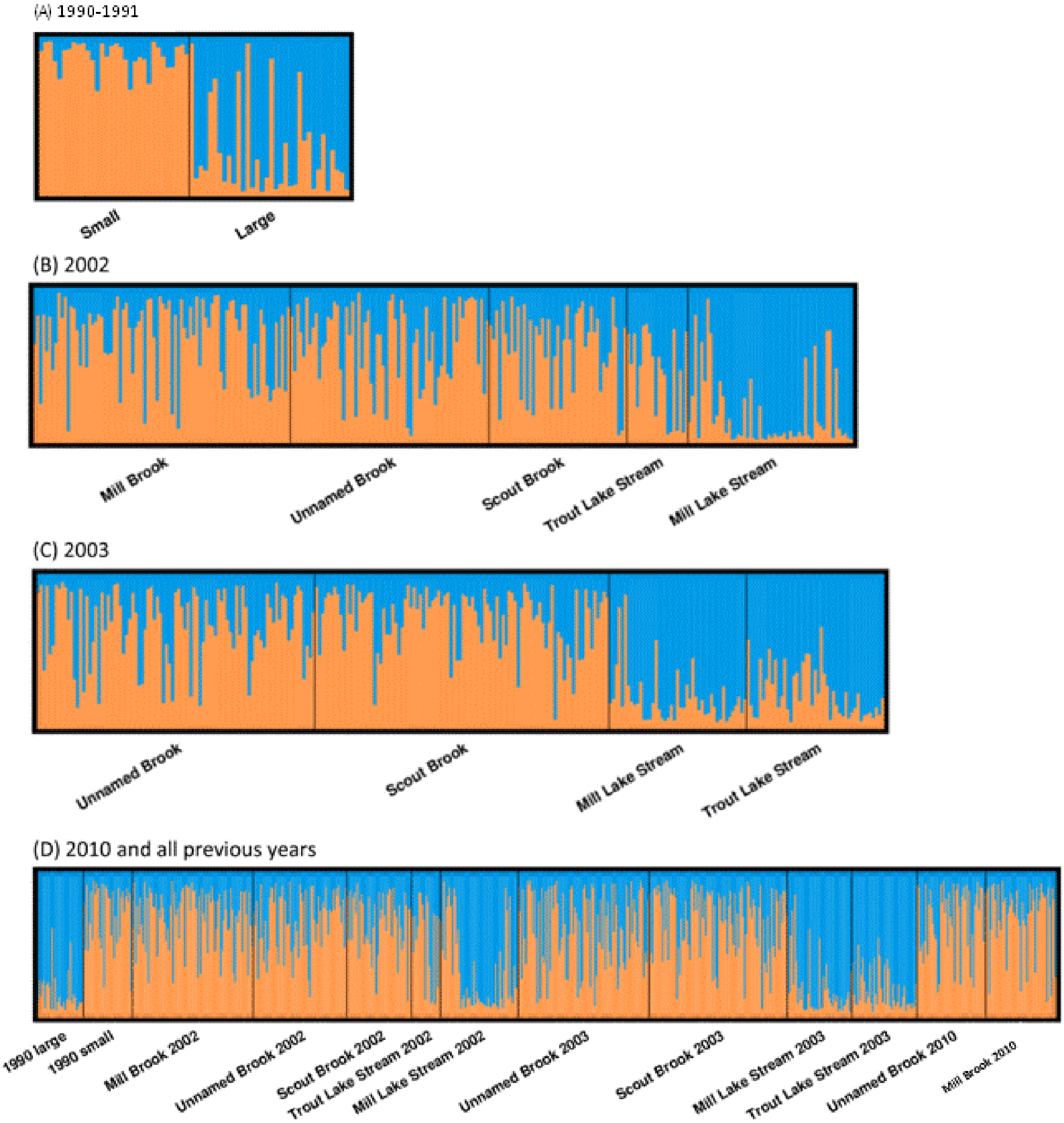


Figure 11. Results of Bayesian clustering in STRUCTURE. L(P) values for number of populations ranging from 1-6 and 250,000 iterations burn in with 500,000 after burn in and three replicates. Delta K calculated following Evanno et al. (2005).



(A) 1990-1991

(B) 2002

(C) 2003

(D) 2010 and all previous years

Figure 12. Results of Bayesian clustering in STRUCTURE for 1990 (A), 2002 (B), 2003 (C), and 2010 with all samples combined (D). See methods for details on STRUCTURE analysis. See Table 1 and Figure 1 for sample details.

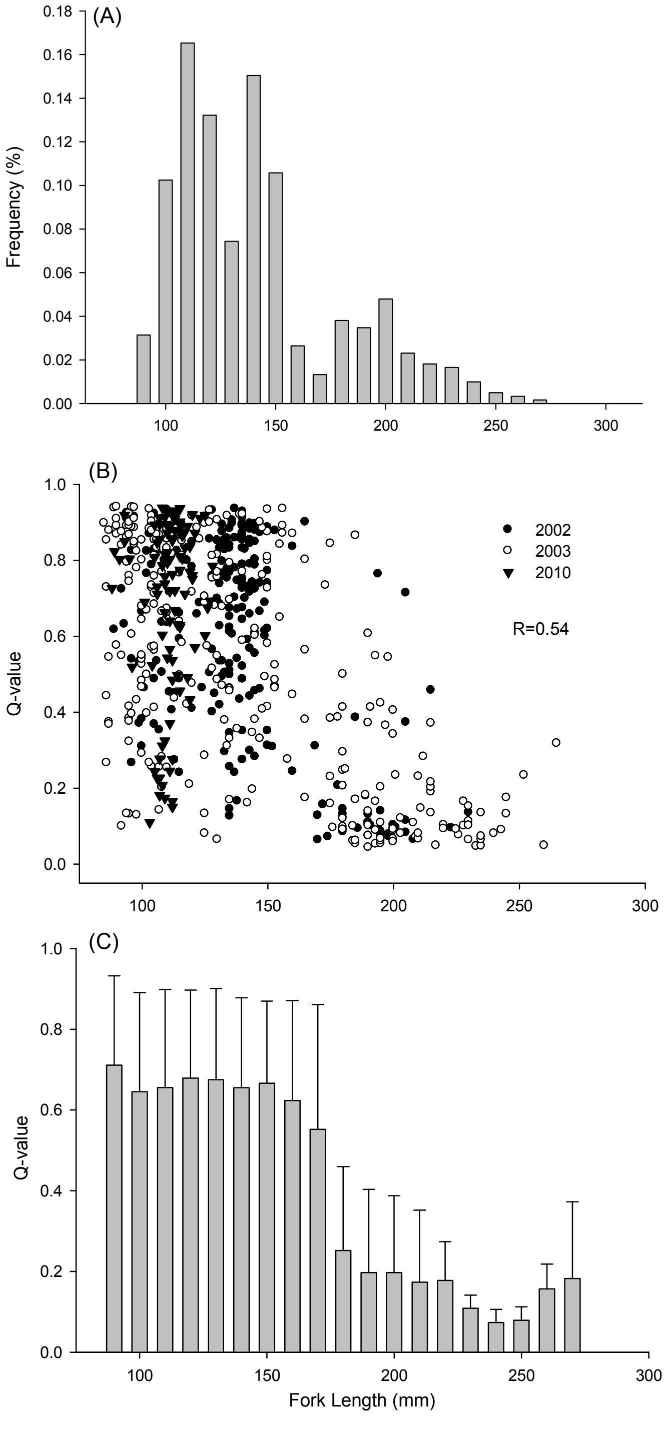


Figure 13. (A) Length frequency distribution of smelt sampled 2002-2010 (B) Q-value from global STRUCTURE analysis and (C) the average Q-value per size class.

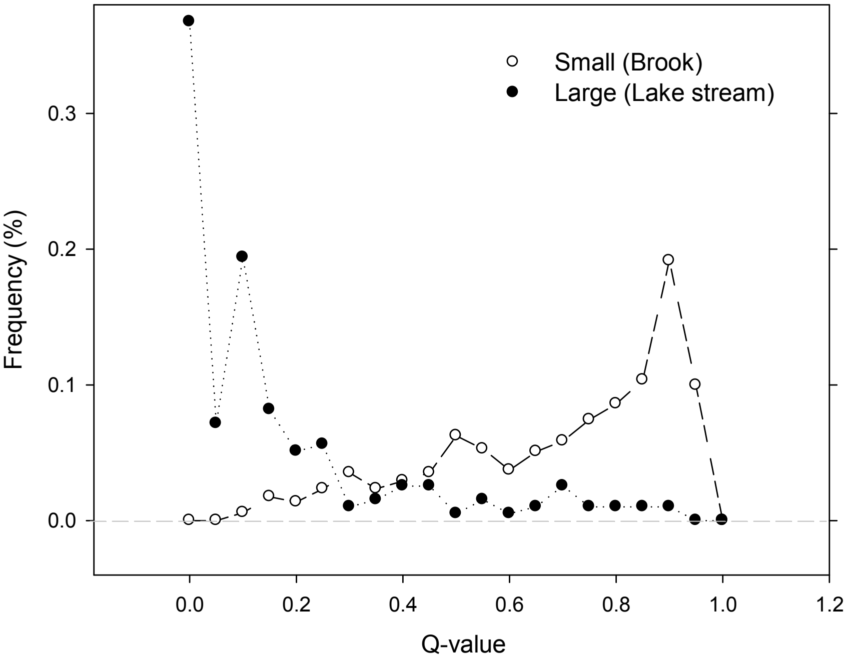


Figure 14. Frequency distribution of Q-values from the STRUCTURE analysis for the two size classes (i.e., >170 mm, and < 170 mm). See Figure 11 and 12 for details regarding Q-value distribution and individual plots.

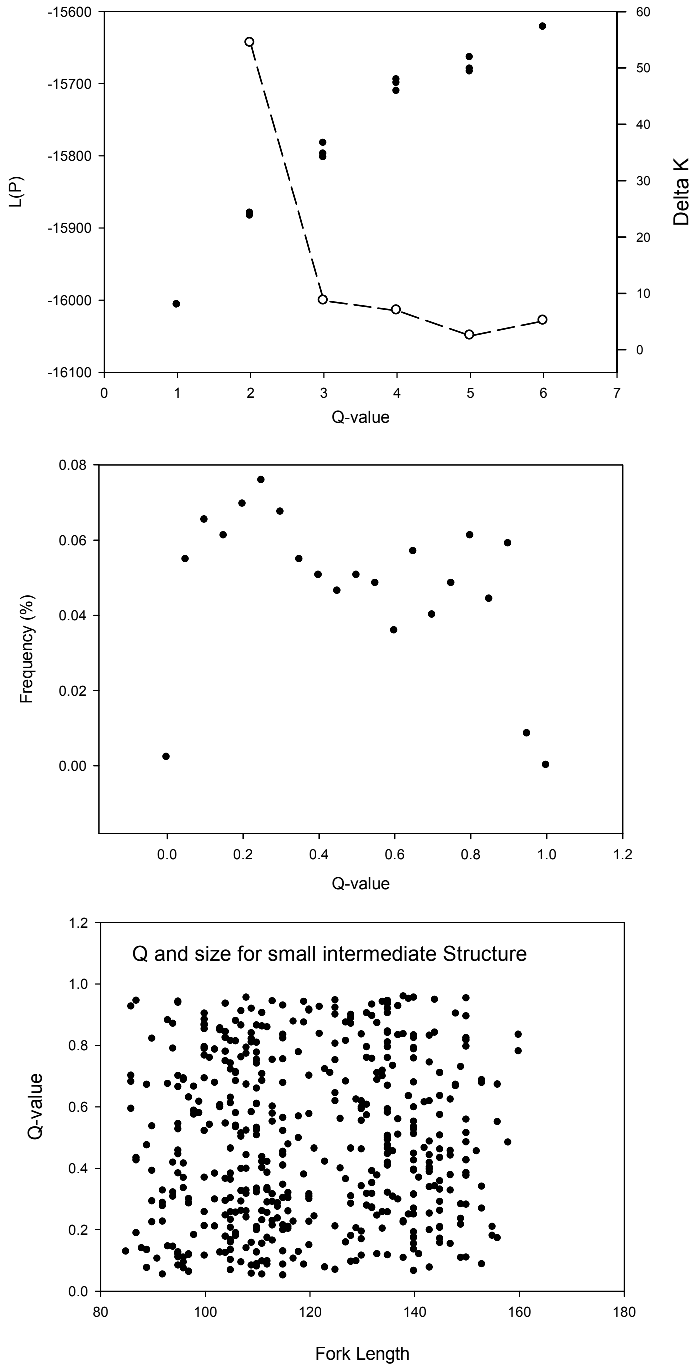


Figure 15. Results of STRUCTURE analysis of all fish <170 mm in fork length. STRUCTURE analysis parameters identical to previous analyses.