

# GREAT LAKES FISHERY COMMISSION

## 2005 Project Completion Report<sup>1</sup>

### COORDINATION OF LAKE SUPERIOR BROOK TROUT GENETIC RESEARCH

By

Chris Wilson<sup>1</sup>

Wendylee Stott<sup>2</sup>

Loren Miller<sup>3</sup>

Brian Sloss<sup>4</sup>

<sup>1</sup>Ontario Ministry of Natural Resources, Trent University, 1600 West Bank Drive, Peterborough, ON  
Canada K9J 8N8 [[chris.wilson@mnr.gov.on.ca](mailto:chris.wilson@mnr.gov.on.ca)]

<sup>2</sup>US Geological Survey, Great Lakes Science Center, 1451 Green Rd., Ann Arbor, MI 48105  
[[wstott@usgs.gov](mailto:wstott@usgs.gov)]

<sup>3</sup>Department of Fisheries, Wildlife & Conservation Biology, University of Minnesota, 1980 Folwell Ave.,  
St. Paul, MN 55108 [[Imm@umn.edu](mailto:Imm@umn.edu)]

<sup>4</sup>Wisconsin Cooperative Fisheries Research Unit, University of Wisconsin-Stevens Point, Stevens Point, WI  
54481 [[Brian.Sloss@uwsp.edu](mailto:Brian.Sloss@uwsp.edu)]

December 2005

<sup>1</sup>Project completion reports of Commission-sponsored research are made available to the Commission's Cooperators in the interest of rapid dissemination of information that may be useful in Great Lakes fishery management, research, or administration. The reader should be aware that project completion reports have not been through a peer-review process and that sponsorship of the project by the Commission does not necessarily imply that the findings or conclusions are endorsed by the Commission. Do not cite findings without permission of the author

## **ABSTRACT**

Parallel, multi-jurisdictional genetic research on the potadromous "coaster" ecotype of brook trout (*Salvelinus fontinalis* Mitchill) in Lake Superior has sparked the need for data sharing and standardization among research labs and management agencies. Marker standardization among collaborating labs using shared reference samples has facilitated standardization of genotype scoring and interpretation, as well as developing a shared toolkit for assessing genetic structure and diversity. This effort has been complemented by the development of a lake-wide genetic database for data sharing among labs, greatly enhancing the resolving power of localized studies as well as lake-wide meta-analyses. Although further steps will be required to fully standardize data collection and interpretation among participating labs, the resultant database of standardized genetic data will facilitate monitoring and adaptive management efforts. Similar initiatives on other species of interest in the Great Lakes basin would benefit all parties (research labs, funding agencies, and management agencies) and should be strongly encouraged.

## **PROJECT BACKGROUND AND RATIONALE:**

Brook trout stocks and fisheries in Lake Superior have declined considerably over the past century (Newman and DuBois 1997). This is of special concern, as brook trout are considered an indicator species of ecosystem health as well as a highly valued component of the native fish community (Newman and DuBois 1997, Newman et al. 1998). Rehabilitating brook trout stocks within the Lake Superior basin has become a priority for both management agencies and non-government organizations (Newman et al. 1998, Schuldt et al. 2003).

Coaster brook trout have warranted particular notice, as these large brook trout appear to be endemic to Lake Superior. As well as their large size, their novel potadromous life history makes them virtually unique among North American salmonines (Behnke 1994). Despite their high profile among management agencies and angler groups, however, it has only recently been resolved that coasters are life history variants or ecomorphs that originate within tributary river populations rather than a distinct genetic group [i.e. evolutionary significant unit (Waples 1995)] (D'Amelio 2002, D'Amelio and Wilson in review). Today, coaster brook trout have been reduced from their former range throughout Lake Superior coastal waters to a handful of extant populations (Newman and DuBois 1997). The urgency of dealing with this drastic reduction has prompted the US Fish and Wildlife Service to develop a coaster broodstock for rehabilitative stocking, despite the general lack of information on basic coaster biology.

Very little is known about the stock structure of brook trout within the Lake Superior basin (Wilson et al. in review). Brook trout are most often identified as belonging to local rivers and bays (e.g. Cypress River, Nipigon River populations, specific watercourses on Isle Royale). Conversely, it has been suggested that all brook trout within Lake Superior and its

tributaries, including Lake Nipigon, constitute one metapopulation (L. Newman, pers. comm.). Resolving the stock structure and spatial scale of gene flow among brook trout populations within the basin is pivotal for implementing management and conservation strategies (Newman et al. 1998, Burnham-Curtis 2000).

Molecular genetic markers are increasingly recognized as valuable tools for stock identification in fisheries management (Carvalho and Hauser 1994, Park and Moran 1994). Information on genetic relationships among populations and taxa of concern is widely recognized as a critical element of conservation and management efforts (Bernatchez 1995), particularly among northern species where stock structure can be difficult to detect using conventional methods (Bernatchez and Wilson 1998). Genetic marker systems have successfully been used to describe geographic genetic structure of brook trout on multiple spatial scales (Angers et al. 1995, Danzmann et al. 1998, Castric et al. 2001, D'Amelio and Wilson, in review). Phylogeographic data from across the species range revealed several mitochondrial DNA (mtDNA) lineages within the Lake Superior basin (Danzmann et al. 1998). Microsatellite DNA markers have been developed specifically to assess microgeographic diversity and stock structure of brook trout (Angers et al. 1995, King and Burnham-Curtis unpubl. data). These have proven highly effective in resolving relationships among brook trout populations on local and regional geographic scales (Angers et al. 1995, Castric et al. 2001, D'Amelio and Wilson in review).

Despite the availability of high-resolution genetic markers, however, significant problems remain in comparing results among genetic studies. Without standardized data collection in terms of shared samples, loci examined, or allelic nomenclature, comparisons between studies are essentially limited to indirect comparisons of diversity (allelic richness

and heterozygosity) and divergence (genetic distance, *F*-statistics, etc.) estimates. This severely limits the comparative power of individual analyses (Paetkau 2003) and is particularly limiting for mobile species or populations that occupy multiple management jurisdictions (Carvalho and Hauser 1994).

Currently, several research groups are investigating different facets of the genetic structure and diversity of brook trout in the Lake Superior basin. The Ontario Ministry of Natural Resources (OMNR) Fisheries Genetics lab has examined the stock structure and identity of coaster and riverine brook trout in Nipigon Bay, Ontario (D'Amelio 2002; D'Amelio and Wilson, in review). This research has shown that coasters within Nipigon Bay are not genetically distinct from river populations of brook trout, and that metapopulation structure exists within Nipigon Bay, with several rivers contributing to the local coaster 'population'. The population genetics lab at the Great Lakes Science Center (GLSC) has similarly investigated the genetic composition of coaster brook trout from Isle Royale and neighbouring waters, and is following up a lake-wide survey of mitochondrial DNA variation by Burnham-Curtis (1996, 2000). Other research efforts have examined the role and effectiveness of existing hatchery stocks in husbanding the genetic resources of wild coaster populations (Cooper 2004), and are focusing on genetic tracking of local restoration efforts (Sloss et al. 2004).

These studies are pursuing largely complementary research questions, using similar genetic marker systems (mitochondrial and microsatellite DNA), and are in various stages of completion. By establishing commonalities among the genetic tools employed and standardizing data scoring, results from these studies can be directly compared and contribute to broader issues within the lake basin. In addition, exchanging data storage models and

establishing a web-accessible database will help ensure that the broader scientific and management community within the Great Lakes basin will also benefit from these studies.

Data management through well-designed relational databases has been identified as a crucial component of both resource management and biological research (Hale 1999, Cole 2000). However, proper database design and management is frequently missing from many research proposals or management plans, to the detriment of many projects (Hale 2000, Paetkau 2003). This is a growing problem in many labs as high-throughput data generation platforms become more accessible; many research labs rely on cobbled-together data storage methods that cannot keep up with the new flood of information.

Two models for a genetic database were considered for this project. The OMNR Fisheries Genetics lab had developed an Access-based database for linking genetic and spatial information for resource management, as part of the Natural Resources DNA Profiling and Forensics Centre at Trent University. This database was designed for tracking individual samples and profiles, from receiving from field sources to final multilocus genetic profiling and assignment to populations. The population genetics laboratory of the Great Lakes Science Center (GLSC)-United States Geological Survey (USGS) had similarly been developing a genetic database to track and manage in-house genetic data, as well as genotype data that was shared with other laboratories. The GLSC database was designed to be implemented in Oracle, and to incorporate USGS and USFW sampling programs in the Great Lakes. The decision was made to go with the GLSC database, due to its more efficient design, better documentation, and substantially greater potential for working with large amounts of data.

The project's objectives were to (1) coordinate research efforts on brook trout genetics in the Lake Superior basin; (2) exchange reference standards and population samples among labs; (3) standardize data collection methods and interpretation (scoring) among labs to facilitate data exchange and broader usage; and (4) develop a genetic database to facilitate data collection, standardization, and retrieval. These objectives addressed identified research priorities by coordinating and encouraging genetic research on brook trout stock structure within the Lake Superior basin (Taylor 1995), and support the Lake Superior brook trout rehabilitation plan (Newman et al. 1998) as well as the GLFC Joint Strategic Plan and Strategic Vision goals.

## **PROJECT COMPONENTS**

### ***DATABASE***

To calibrate the genetic profiling results from the different labs, Wendy Stott and personnel at the USGS Great Lakes Science Centre (GLSC) developed a database specifically for managing genetic data for fish species in the Great Lakes. To manage this volume of data efficiently, the database tracks field collections, sample inventories, genetic methodologies, and the resultant genetic data. The database meets standards for metadata suggested by the National Biological Information Infrastructure program of the United States Geologic Survey (<http://www.nbio.gov/datainfo/metadata/standards/>). Metadata is a description of characteristics of data such as quality, provenance, and condition. The description of the data is organized in a standardized format that can ease the search for information. Metadata ultimately makes information about data sets more easily accessible to scientists and researchers.

The database is broadly applicable for genetic data storage and retrieval for a variety of species and/or genetic marker and analysis systems. An application built with the database provides standard data summaries and future applications (currently planned to be completed by the end of 2006) will facilitate publication of information on the World Wide Web.

Entities in the database can be broken down into seven main components, from initial field collections to sample processing results of genetic analyses (Figure 1). The collection module describes a set of one or more individuals collected in the same manner from the same geographical location, possibly at the same time. As well as recording the geographic location of the sample collection, the database accommodates information on capture method, gear used, and collector(s). Organism-related tables track information on taxon identity (e.g. known or suspected species or hybrid), life history stage and other biological attributes, field markings if any (e.g. clip marks), and population or stock. In-house tracking is facilitated by components or modules dealing with sample management and storage. Beginning with DNA preservation method, extraction, and storage, the data flow addresses the genetic analyses employed, with several hierarchical levels of detail possible depending on how much specific detail is desired by users. The analysis results tables can store a variety of genetic data, ranging from allozyme, microsatellite, sequence or RFLP data.

An important feature of the database is its ability to accommodate data from outside labs (separate users) in order to calibrate shared data among labs. As well as tracking sample exchange among labs, the GLSC database can import external genetic data, facilitating meta-analyses of pooled data. In order to correct for potential scoring differences among separate labs, the database can also correct for systemic scoring differences between labs as a query function. This is a key feature for sharing collaborative projects among multiple labs, as

different labs may produce slightly different genotype scores depending on lab equipment, amplification conditions, visualization platform, and scoring rules (LaHood et al. 2002, Koumi et al. 2004, Moran et al. 2005, this study).

The logical and data models for the GLSC database have been completed, and several modules are currently functional. All the data from the brook trout analyzed by the various agencies involved in this project have been stored in the database at the Great Lakes Science Center and data from any future contributors will also be deposited there. The database will also be populated with the genetic results of the component Lake Superior studies following their completion and publication. Once all modules have been compiled, the database will be beta-tested before it is made available on the Internet. Collaborating labs, partner agencies, and members of the public will have differing levels of access, with access to proprietary data controlled by secure login. Once published, the database architecture itself will be freely available on request (W. Stott, pers. comm.).

### ***GENETIC DATA COLLECTION AND ANALYSIS***

Initial standardization efforts involved sharing DNA from wild and hatchery brook trout among labs in order to directly compare results (Figure 2). As wild coaster brook trout are in limited supply and high demand, however, the decision was made to rely on hatchery samples for data standardization. Brook trout from the OMNR Lake Nipigon broodstock were therefore selected as the reference set for comparison among labs, as this strain of fish has been used by several agencies in different parts of the Lake Superior basin for coaster restoration efforts (Newman and DuBois 1997).

Tissue samples from the 1999 broodstock yearclass of the Nipigon brook trout hatchery strain (OMNR Fish Culture broodstock lot 99DN-126) were obtained from the OMNR Fisheries Research Facility (Codrington, ON). All fish were individually branded and PIT-tagged to enable resampling or selective breeding for potential future use for establishing living reference standards. Nonlethal samples were taken via hole punch of the caudal fin and individually preserved in 95% ethanol. Genomic DNA was extracted at the OMNR Fisheries Genetics lab and subdivided into replicate aliquots which were shipped to participating labs along with approximately 40 mg of surplus tissue, providing equal original source tissue and DNA for parallel genotyping. By having one lab extract the genomic DNA and sharing identical sets of subsamples among participating labs, this study eliminated potential variation due to differences in DNA extraction method or tissue quality, and focused specifically on variation or differences due to different amplification methods, instrumentation (sequencers) and interpretation among labs.

During the GLFC-sponsored coaster brook trout workshop in 2003, the project was discussed with other genetic researchers, and Dr. Brian Sloss (U. Wisconsin / Wisconsin DNR) was recruited to the project. This strengthened the project considerably, bringing the number of collaborating labs up to four and including a researcher with no previous working experience with the species.

Genetic profiling of individual fish employed microsatellite loci from two microsatellite libraries developed specifically for brook trout (Table 1), and included both published and unpublished loci (Angers et al. 1995; Tim King, USGS, Leetown, WV unpubl. data). A total of 19 loci were used (Table 1), although not all loci were used within each lab (Table 2). The loci showed very different phenotypes when visualized via electrophoresis,

with dinucleotide loci showing much stronger stutter patterns (amplification artefacts) than tri- or tetranucleotide loci (Figure 3). As the King library mostly consists of microsatellites with tri- and tetranucleotide core repeats, it was expected that they might show lower variation among labs than those developed by Angers et al. (1995).

Allele sizes for each individual and locus were obtained and scored independently by each lab, using a range of visualization platforms. The GLSC and OMNR genetics labs used in-house automated sequencers (ABI 310 and 3100 capillary sequencers, and an ABI 377 (slab gel) sequencer, respectively), whereas both the UMN and UWI labs relied on core sequencing facilities that employed ABI 377 slab gel sequencers. The latter labs received binned data from their respective core facilities, whereas the GLSC and OMNR labs directly obtained raw data which was subsequently binned. Since the database was developed and housed at the GLSC genetics lab, this lab was used as the reference dataset for comparison among labs.

Comparison of initial scoring among labs showed strong congruence in scoring individual genotypes as homozygous or heterozygous (Figures 4 and 5). The most similar datasets were those from the GLSC and OMNR labs, which shared the greatest number of loci (Figure 5). Despite the more "noisy" phenotypes of dinucleotide loci, there was no clear pattern of scoring differences with core repeat size. Results for loci from the older dinucleotide microsatellite library (Angers et al. 1995) showed strong correlations among labs (Figures 4 and 5), whereas several of the unpublished loci (*SfoB52*, *SfoC153*, and *SfoD105*) showed substantial scoring discrepancies among labs (Figure 6). Excluding these loci resulted in approximately 90 percent overall concordance among labs without rescoring of allele sizes.

The largest source of different scoring among labs were offsets or consistent differences in scored allele sizes (Figure 6, Table 2). These differences may be caused by different amplification conditions (e.g. "allele plus A" artefacts resulting from high magnesium ion concentration and/or surplus *Taq* polymerase in PCR reactions; Fishback et al. 1999), but are more often due to different electrophoretic conditions or machine platforms (Haberle and Tautz 1999, Koumi et al. 2004, Moran et al. 2005), or different binning rules employed by labs or core sequencing facilities. These systemic differences are readily corrected once recognized, and do not pose a serious impediment to merging or standardizing data. Size offsets varied among loci as well as among labs (Table 2), consistent with results reported elsewhere (Delmotte et al. 2001). As three of the four labs used ABI 377 automated sequencers, machine platform was not a probable source of the observed differences. Rather, the locus-specific differences in lab offsets are suggestive that lab-specific scoring and binning rules were the major sources of discrepancies among labs. These can be readily corrected by discussion and consensus establishment of scoring and binning rules, to be shared among participating labs.

Figure 6 shows the recorded comparative scoring of alleles from individual brook trout, sorted by allele size and using the GLSC data as the baseline for comparison, prior to comparison and consultation among labs with resultant secondary proofreading and corrections. Differences in scoring among labs are emphasized by the differing vertical axes for each locus, which reflects both size shifts among alleles at a locus and the number of alleles at that locus. For the most part, labs showed good agreement in their interpretation of allele sizes. Although one of the original intents of the calibration project was for all three original participant labs to standardize all available loci, this was not achieved, as the

University of Minnesota lab was unable to screen all loci due to personnel constraints.

Bringing a fourth lab (Brian Sloss, University of Wisconsin) into the study, however, made up for this original shortfall and provided additional resolution for the eight shared loci (Figure 6).

Correcting for the observed offsets resulted in very consistent genotyping among labs, showing that datasets can be merged with confidence upon standardization. Despite strong overall agreement, however, some loci showed large differences in scored alleles among labs (Figure 6). Rescoring of sequencer output confirmed that some of these differences were real, and may have been caused by PCR artifacts (reamplification of problematic samples), whereas others were caused by scoring and binning errors. Disagreement among labs for data at several loci (*Sfo23*, *SfoC113*, *SfoD75*) was confirmed by re-scoring of the original data, but was identified as being due to loading too much DNA into the OMNR sequencer (C. Wilson, pers. obs.). This was subsequently resolved by diluting and re-analysing the amplified DNA, which resulted in scores more consistent with those from the other labs (data not shown).

Of the 19 loci employed in this study, 15 are good candidates for inclusion in a standardized set of microsatellite loci for brook trout genetic studies. Of these, 9 loci (*Sfo8*, *Sfo12*, *Sfo23*, *SfoC28*, *SfoC79*, *SfoC86*, *SfoC88*, *SfoC115*, *SfoC129*) were highly consistent with low error rates among labs (Figure 6). An additional six loci (*Sfo18*, *SfoC24*, *SfoC38*, *SfoC113*, *SfoD75*, and *SfoD100*) also showed strong concordance after initial data comparison and correction, including re-analysis of diluted PCR product as outlined above. Three loci (*SfoB52*, *SfoC153* and *SfoD105*; T. King unpubl.) failed to produce consistent results among labs and would not be good candidates for establishing a standardized genetic analysis toolkit, although discrepancies at loci *SfoB52* and *SfoC153* could potentially be resolved if allelic

ladders (mix of reference alleles of known sizes) were available. An additional locus (*SfoD9.1*; T. King unpubl.) showed strong evidence of a null allele (problematic amplification and strong heterozygote excess), and was dropped from the analysis. For the time being at least, the use of these four loci is not recommended for brook trout genetic studies.

### *Sources of variation*

As described above, a number of sources contributed to interpretative differences among the participating labs. These included:

- differences in interpreted allele sizes among labs. The high allelic diversity for microsatellite DNA loci in most species and populations, and relatively small percentages of populations surveyed (excluding humans and domesticated animals), precludes a modal or alphabetical nomenclature. Instead, the most reliable scoring is based on interpreted allele sizes (length in nucleotides) as inferred from electrophoretic outputs alongside known size standards. These inferred allele sizes relative to known standards are analogous to  $R_f$  values (relative electrophoretic mobility scores) for isozyme data (Shaklee et al. 1990). Unfortunately, the different designs of slab gel versus capillary electrophoresis sequencers can result in mobility shifts for identical alleles between automated sequencer platforms, even in comparison with size standards (Koumi et al. 2004, Moran et al. 2005). In a similar but much larger standardization effort, Moran et al. (2005) documented significant differences in allele size scoring among labs but avoided scoring conflicts by (a) working from a dataset of known (reference) allele sizes and (b) using modal scoring (i.e. smallest to largest alleles) for initial interpretation before converting back to true allele sizes. This study did not have the advantage of a curated set of loci with known allele sizes; indeed, one of the primary goals of the study was to establish these.
- user experience and familiarity with specific loci. The oldest and most established loci (*Sfo8*, *Sfo12*, *Sfo18*, and *Sfo23*; Angers et al. 1995) have been widely used and have generated enough familiarity among users to promote consistent scoring and

interpretation, despite having more stutter and artefact bands than the unpublished King loci (Figure 3). In several cases, amplification and scoring were performed by undergraduate helpers (L. Miller, pers. comm.), which may have led to scoring or binning errors due to lack of experience. Data proofreading and correction is made more difficult when core facilities provide binned data only, as was the case for two of the participating labs (Miller and Sloss). These differences highlight the importance of archiving raw output from data generating platforms, rather than relying solely on binned (i.e. user-interpreted) data (see Appendices I and II).

- non-template adenylation of alleles (addition of extra adenine nucleotides, or “allele+A”) by *Taq* polymerase during PCR amplification of microsatellite loci (Fishback et al. 1999). This can be compensated for by adding post-amplification extension times to PCR reactions, addition of surplus *Taq*, or altering amplification conditions (Fishback et al. 1999; Pompanon et al. 2005).
- binning errors. This was a significant source of interpretation differences among labs, resulting from differences in converting raw sequencer output scores to integer values. Automated sequencers are very precise in sizing alleles, but not necessarily accurate (Haberle and Tautz 1999, Koumi et al. 2004, Pompanon et al. 2005). Alleles are sized against size standards (DNA fragments of known size) which are included in each sequencer run, with allele sizes estimated via curvilinear regression. Although true allele sizes are integers (e.g. 100 vs 104 nucleotides long), their estimated sizes tend not to be (e.g. 101.55). Problems occur when trying to sort out what size class an estimated value truly belongs in (e.g. whether an allele of estimated length of 101.55 nucleotides represents a 100 or a 104 allele).

Several sequencers now come with integral software for binning microsatellite data based on multimodal distribution of interpreted allele sizes (Idury and Cardon 1997). Another approach is to directly or manually examine the distribution of estimated allele sizes from the raw sequencer output, and infer allele sizes based on frequency peaks corresponding to size differences that match multiples of the core repeat sequence for that locus (e.g. 150, 152, 154 for a dinucleotide locus; 160, 163, 166, etc. for a trinucleotide locus). Binning rules are often lab-specific, and may vary depending on the person scoring the data. It is therefore advisable for data to be scored by at least two people, and preferably done automatically by software which is included with most sequencers, followed up by manual proofreading by two or more users (see QA recommendations in Appendix I).

These issues can be avoided by providing the raw sequencer data to a central repository (in this case the GLSC database) so that offsets can be determined with a minimum of filtering or interpretation by different users (see recommended requirements for lab participation in standardization efforts, Appendix II).

Moran et al. (2005) largely avoided this problem by combining a modal scoring system (e.g. rank order, smallest to largest alleles) with a reference set of alleles and genotypes where true allele sizes were known in advance. Establishing similar allelic ladders for brook trout loci would greatly improve scoring consistency among labs.

- loading too much PCR product for visualization (e.g. *SfoC113*, *SfoD75*, OMNR lab).  
Overloading an excess of amplified DNA into an automated sequencer can result in electrophoresis artefacts such as visualization of non-allelic PCR artefacts or increased mobility (smaller apparent sizes) of microsatellite alleles (Fernando et al. 2001). This can

be readily corrected by dilution of PCR products to the point that artefact bands disappear but genotypes are still readily detectable (typically 1:5 to 1:10 dilution of PCR product in water) (Fernando et al. 2001).

These sources of error are increasingly being recognized in genetic studies, as well as their consequences and recommended solutions (summarized in Bonin et al. 2004, Pompanon et al. 2005). These can largely be prevented or corrected using well-documented protocols and quality assurance procedures (outlined in Appendix I). Data quality control is particularly important where data are shared among labs for collaborative purposes or standardization efforts; suggested criteria for participation in standardization efforts are listed in Appendix II.

Despite the discrepancies observed among labs, overall consistency was quite high. Figure 7 summarizes scoring differences and consistency among labs before subsequent data comparisons and corrections. Of the 19 loci considered, 13 showed mean concordance among labs of 80% or more before data correction. There was a general pattern of reduced agreement in scoring with increasing allelic diversity, although two loci (*Sfo18* and *SfoC153*) showed poor initial agreement despite having relatively few alleles (Figure 7). In general, the unpublished loci from Tim King showed better concordance among labs than the older markers published by Angers et al. (1995), but this apparent discrepancy disappeared after data comparison and correction among labs, with the majority of loci showing more than 95% concordance among labs.

Each lab showed some unique variation at particular loci, but data quality and compatibility was very good. The comparison and offset corrections among labs highlighted the importance of data standardization. The observed offsets in scored allele sizes among the

different labs, if left uncorrected, would have crippled any meta-analyses that relied on allele-specific information. By correcting for scoring differences among labs, however, all genetic data on brook trout from within each participating lab can now be standardized and used to resolve lake-wide genetic structure of brook trout in Lake Superior. By establishing these common data standards and establishing links among formerly disparate datasets, this study has thereby enhanced the value of both existing genetic data and ongoing research for brook trout rehabilitation and conservation efforts.

### ***FUTURE DIRECTIONS***

As stated earlier, data management is a constant challenge for many genetic labs, particularly with rapidly evolving technologies and personnel turnover. It can reasonably be expected that technological advances and corresponding throughput capability will continue to grow, making data management an ever-increasing challenge. At the same time, shrinking research and management budgets will encourage data exchange among research labs and management agencies. The genetic database developed in conjunction with this project can be used to prevent data loss and promote greater data continuity within labs as well as fostering data sharing among labs and management agencies.

Since the completion of the data collection and standardization component of the project, two additional labs (Kim Scribner, Michigan State University; Barbara Evans, Lake Superior State University) have asked to participate in the brook trout genetic data standardization effort. These labs have been provided with replicate DNA and tissue sets from the reference set of brook trout, as well as information on loci used and amplification conditions. The data from these labs have not yet been incorporated into the GLSC database,

and will be used to beta-test the data entry protocols and participation criteria (Appendices I and II).

To standardize scoring of chinook salmon microsatellite genotypes and alleles, Moran et al. (2005) developed documentation and scoring rules for each locus. Each document describes the source paper where the locus primer sequences were originally published, primer and core repeat sequences, amplification conditions and scoring rules. Scoring instructions include example screen-capture images of typical allele phenotypes and known artefacts, as well as contact details for the curating lab(s) and individuals. Based on this example, we are working to develop similar documentation for the brook trout loci tested in this study, which should simplify interpretation and participation by other labs.

As a result of this study, data can now be directly compared among labs which employ the reference standards and evaluated microsatellite loci. This will permit quantification of differentiation among studied brook trout stocks, and provide baseline data for determining the location, status and spatial extent of additional stocks as well as examining the consequences of fish stocking and land/water uses on the genetic diversity and fitness of wild fish. Establishing common data standards among the different genetics labs pursuing brook trout research in Lake Superior will enhance the value of existing data, help integrate independent efforts, and build a lake-wide genetic database for brook trout rehabilitation and conservation efforts. Although more work needs to be done to simplify and standardize data collection and interpretation among labs, considerable progress has been made. Implementing quality assurance and controls within and among participating labs will help ensure data quality and applicability, and facilitate analyses of harmonized datasets.

The first application of this project has been to standardize the existing GLSC and OMNR microsatellite datasets for a preliminary meta-analysis of lake-wide data from coaster brook trout (Figure 8; Stott et al. 2004). The analysis indicates that brook trout in Lake Superior are grouped into at least five regional metapopulations, and that coaster brook trout from Isle Royale and Nipigon Bay are clearly distinct despite their geographic proximity (Figure 8). These results are encouraging, as they indicate that substantial genetic structure and variation still persists among remnant coaster populations in the lake.

The results of this study were presented at the genetic data standardization symposium held at the 2004 American Fisheries Society conference (Wilson et al. 2004), and at the Great Lakes Regional Data Exchange (Stott et al. 2004). Discussions with researchers involved in similar projects reinforced the value of standardized data collection and scoring, and suggested that this project avoided some significant pitfalls. By actively communicating and sharing samples among labs, the project was not subject to personality conflicts or sample hoarding, as was reported for other standardization projects. Also, establishing data storage and retrieval up front bypassed significant problems encountered elsewhere. Data ownership and access rights have been a major impediment elsewhere (P. Moran, pers. comm.; W. Ardren, pers. comm.), but can be circumvented by agreements between labs and management agencies regarding ownership, access, publication timelines, and sunset clauses.

## **RECOMMENDATIONS**

Based on the results of this study and their demonstrated utility, it is recommended that standardized set of loci or other genetic markers be established for economically and ecologically significant species. Although data and locus sharing goes on informally among

research labs, all participants would benefit from a core set of standardized tools that can readily be applied to local or basin-wide research or management issues. This could be facilitated by small grants specifically focused on developing, implementing, and/or documenting available genetic tools, and establishing a standardization kit for the taxon and marker systems. This would ideally be comprised of documentation of genetic markers, including primer sequences and labels, amplification conditions, and scoring rules, and sets of reference DNA to be used for standardization. This reference DNA set could consist of either reference samples from geographically diverse source populations, as archived tissue or living specimens, or locus-specific allelic ladders. Although one of the primary goals of the study was to establish a reference set of genotypes to facilitate standardized genetic data collection, calibration of scoring differences among labs was considerably hindered by not having advance knowledge of true allele sizes or genotypes. It is therefore strongly recommended that reference standards with confirmed genotypes and/or allelic ladders be established for species of interest, in order to facilitate standardization efforts.

To maximize the value of funded genetic research on Great Lakes species, the GLFC should require that raw data from each study be placed in a centralized data repository, or that other researchers doing GLFC-funded research on the same species be allowed access to the data on request. This would likely generate issues over data sharing and ownership, but would be of enormous value for tracking species and populations both spatially and temporally.

Benefits to funded projects would include substantially reduced startup times and costs due to an established core set of genetic markers, simplified scoring and interpretation due to comparison with established standards, improved quality assurance and control based on

reference standards. Ironically, parallel (repeated) analysis of reference standards by each participating lab will reduce overall cost and effort to labs and funding agencies, by reducing the need for field sampling and parallel analysis of wild or hatchery populations. Funding and management agencies would thereby effectively leveraging research dollars spent in addition to benefiting from implemented QA/QC protocols. Individual projects and management agencies would also benefit by being able to draw on information from other studies or jurisdictions, thereby enhancing the value and power of the specific project and increasing its potential application to the Great Lakes basin as a whole.

## **ACKNOWLEDGEMENTS**

This work was funded by the Great Lakes Fishery Commission Coordination Activities Program. Dave Oliver (Skylark Information Services) provided essential support in developing the database logical model and architecture. Tissue samples for comparison among labs were provided by Bill Sloan (Codrington Fisheries Research Facility, OMNR). Peter Addison (OMNR), Mohamad-Ali Kolaitat (GLSC), and several undergraduate students provided significant help with data collection. Feedback and suggestions from Chuck Krueger (GLFC) were instrumental in strengthening the modified completion report.

## REFERENCES

- Angers, B., L. Bernatchez, A. Angers, and L. Desgroseillers. 1995. Specific microsatellite loci for brook charr reveal strong population subdivision on a microgeographic scale. *Journal of Fish Biology* 47 (Suppl. A): 177-185.
- Bernatchez, L. 1995. A role for molecular systematics in defining evolutionary significant units in fishes. p. 114-132 *in* (J.L. Neilson, editor). *Evolution and the aquatic ecosystem: defining unique units in population conservation*. American Fisheries Society Symposium 17, Bethesda MD.
- Bernatchez, L. and C.C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* 7: 431-452.
- Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.
- Brinkmann, B., M. Klitschar, F. Neuhuber, J. Huhne and B. Rolf. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* 62: 1408-1415.
- Burnham-Curtis, M.K. 1996. Mitochondrial DNA variation among Lake Superior brook trout populations: summary of genetic analyses. Research completion report prepared for U.S. Fish and Wildlife Service, Ashland Fisheries Resource Office, Ashland, WI. 29p.
- Burnham-Curtis, M.K. 2000. Genetic profiles of selected brook trout *Salvelinus fontinalis* populations from Lake Superior. Research completion report prepared for U.S. Department of the Interior, Fish and Wildlife Service, Ashland Fisheries Resource Office. 39p.
- Castric, V., F. Bonney, and L. Bernatchez. 2001. Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* 55: 1016-1028.
- Carvalho, G.R. and L. Hauser. 1994. Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* 4: 326-350.
- Cole, K. 2000. Laboratory database to manage electrophoresis and chromatograph separations and the associated samples. *Biotechniques* 29: 1256-1262.
- D'Amelio, S. 2002. Conservation genetics and metapopulation structure of brook trout (*Salvelinus fontinalis*) in Nipigon Bay (Lake Superior, Ontario). Master's thesis, Trent University, Peterborough.
- D'Amelio, S. and C.C. Wilson. Metapopulation structure among source populations for coaster brook trout in Nipigon Bay, Lake Superior. *Transactions of the American Fisheries Society* (in review).

- Danzmann, R.G., R.P. Morgan, M.W. Jones, L. Bernatchez, and P.E. Ihssen. 1998. A major sextet of mitochondrial DNA phylogenetic assemblages extant in eastern North American brook trout (*Salvelinus fontinalis*): distribution and postglacial dispersal patterns. *Canadian Journal of Zoology* 76: 1300-1318.
- Delmotte, F., N. Leterme, and J.-C. Simon. 2001. Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. *BioTechniques* 31: 810-818.
- Fernando, P., B.J. Evans, J.C. Morales, and D.J. Melnick. 2001. Electrophoresis artefacts - a previously unrecognized cause of error in microsatellite analysis. *Molecular Ecology Notes* 1: 325-328.
- Fishback, A.G., R.G. Danzmann, T. Sakamoto, and M.M. Ferguson. 1999. Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout *Oncorhynchus mykiss*. *Aquaculture* 172: 247-254.
- Haberl, M. and D. Tautz. 1999. Comparative allele sizing can produce inaccurate allele size differences for microsatellites. *Molecular Ecology* 8: 1347-1349.
- Hale, S. 1999. How to manage data badly, Part I. *Bulletin of the Ecological Society of America* 80: 246-268.
- Hale, S. 2000. How to manage data badly, Part II. *Bulletin of the Ecological Society of America* 81: 101-103.
- Idury, R.M. and L.R. Cardon. 1997. A simple method for automated allele binning in microsatellite markers. *Genome Research* 7: 1104-1109.
- Koumi, P., H.E. Green, S. Hartley, D. Jordan, S. Lahec, R.J. Livett, K.W. Tsang, and D.M. Ward. 2004. Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* 25: 2227-2241.
- LaHood, E.S., P. Moran, J. Olsen, W.S. Grant, and L.K. Park. 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Molecular Ecology Notes* 2: 187-190.
- Moran, P., T.A. Lundrigan, and D.J. Teel. 2004. Coastwide standardization of chinook salmon microsatellite data for international fixed fishery harvest management. AFS genetic marker standardization symposium, Madison WI, August 2004.
- Moran, P., M. Banks, T. Beacham, C. Garza, S. Narum, M. Powell, M. Campbell, L. Seeb, R. Wilmot, S. Young, W. Ardren, and J. Wenberg. 2005. Interlaboratory standardization of coast-wide chinook salmon genetic data for international harvest management: draft progress report from the Genetic Analysis of Pacific Salmonids (GAPS) consortium to the Chinook technical Committee of the Pacific Salmon Commission. October 2005.

- Newman, L.E., and R.B. DuBois (editors). 1997. Status of brook trout. Lake Superior Technical Committee, Great Lakes Fishery Commission. 33p.
- Newman, L.E., R.B. DuBois, and T.N. Halpern (editors). 1998. A brook trout rehabilitation plan for Lake Superior. Great Lakes Fishery Commission. 28p.
- Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology* 12: 1375-1388.
- Park, L.K. and P.A. Moran. 1994. Developments in molecular techniques in fisheries. *Reviews in Fish Biology and Fisheries* 4: 272-299.
- Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6: 847-859.
- Schuldt, J.A., E. Iwachewski, D. Schreiner, K. Cullis, R.W. Mackereth, C. Huckins, M. Jennings, and C. Wilson. 2003. Coordination of Coaster Brook Trout Rehabilitation Efforts. 2003 AFS meeting, Laval, Quebec, August 2003.
- Schwengel, D.A. , A.E. Jedlicka, E.J. Nanthakumar, J.L. Weber, and R.C. Levitt. 1994. Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics* 22: 46-54.
- Shaklee, J.B., F.W. Allendorf, D.C. Morizot, and G.S. Whitt. 1989. Genetic nomenclature for protein-coding loci in fish: proposed guidelines. *Transactions of the American Fisheries Society* 118: 218-227.
- Stott, W., C. Wilson, L.M. Miller, and B. Sloss. 2004. Standardization of genetic markers and databases. Great Lakes Regional Data Exchange (RDX) conference: Decision Support for the Great Lakes – St. Lawrence System for the New Millennium. Detroit MI, October 2004.
- Waples, R.S. 1995. Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. p. 8-27 *in* (J.L. Neilson, editor). *Evolution and the Aquatic Ecosystem: defining unique units in population conservation*. AFS symposium 17, Bethesda MD.
- Wilson, C., W. Stott, L.M. Miller, and B. Sloss. 2004. Standardization of genetic markers and databases for coaster brook trout in Lake Superior. AFS genetic marker standardization symposium, Madison WI, August 2004.
- Wilson, C.C., W. Stott, L. Miller, S. D'Amelio, M. Jennings, and A. Cooper. Conservation genetics of Lake Superior brook trout: issues, questions and directions. *North American Journal of Fisheries Management* (in review).

**Appendix I:** Quality assurance / quality control recommendations for tracking, assessing and avoiding data collection errors before and after genotyping (modified from Bonin et al. 2004).

<b>Step</b>	<b>Recommendations</b>
<b>Whole process</b>	Perform blind analysis on known and/or replicate samples
<b>Sampling</b>	
Before	Use a standard protocol for labelling and conserving samples, including species, population, individual, sampling date, collector and archival methods
	Record individual sample identifiers and related details in a database
<b>DNA extraction</b>	
	Include negative controls in extractions to monitor for sample cross-contamination
	Follow a rigorous, well-documented extraction protocol
	Extract DNA in a dedicated (preferably separate) area/room from where DNA amplification and/or visualization occurs
<b>DNA treatment/amplification</b>	
Before	Include negative controls to monitor for reagent contamination or sample cross-contamination
	Check in-house data reproducibility (pilot study)
	Follow a rigorous, well-documented amplification protocol
	Discard samples with amplification/repeatability problems
	Include previously typed samples in each amplification as references
After	Carry out replicate amplifications (5-10% of sample set) to confirm genotyping
	Perform independent repeatability tests (collaborative labs)
<b>Scoring</b>	
Before	Obtain a good experience with the marker scoring prior to any genotyping session
During	Use automated scoring where possible; confirm results manually
	Use reference samples with known genotypes to control scoring
After	Have two independent readers score the data and replicate genotypes to confirm data interpretations
<b>Analysis</b>	
Before	Correct mis-scoring or interpretation (binning) errors; eliminate suspicious markers
After	Quantify the overall genotyping error rate; report results before and after data correction

**Appendix II:** Recommendations for conditions to be met by genetic laboratories wishing to participate in genetic standardization projects hosted within the GLSC genetic database.

One purpose of the genetic calibration and standardization for brook trout was to establish a prototype for standardizing and sharing genetic data among laboratories and management jurisdictions, thereby increasing cost efficiencies and enhancing data value for participating labs, management agencies, and funding agencies. As such, it is hoped that additional genetic labs will choose to participate in this and other genetic standardization projects. To ensure the quality of data that are entered into the database, however, several quality standards must be met. The following conditions are therefore suggested as requirements to be met in order for a lab to be included in and have access to the GLSC database:

*Advance requirements*

- 1) demonstrated (published) expertise working with microsatellite data.
- 2) provide evidence (documentation) of data collection and internal QA/QC protocols.

*Genotyping requirements*

- 3) correct genotyping of reference samples:
  - $\geq 95\%$  overall correct scoring of reference genotypes;
  - no scores  $< 90\%$  correct for any single locus (not including missing data);
  - less than 25% missing data overall.

*Data sharing requirements*

- 4) willingness to provide raw and interpreted data to GLSC database. The lab generating the data will retain proprietary ownership, but must be willing to make the data available after publication for inclusion in the Great Lakes dataset.
- 5) provide written permission to allow Great Lakes management agencies which fund the data collection access (prior to the data being formally published) if needed for use in support of achieving management goals for that species. This access would be granted with the provisos that (a) the generating lab would be consulted prior to the data's use, (b) proprietary data would not be subsequently distributed, and (c) baseline data would not be published other than by the lab which collected the data.
- 6) the candidate lab would in turn agree not to publish proprietary data in the GLSC database that was collected by other labs.

These conditions and standards must be agreed to and met before any data will be accepted from the candidate lab for inclusion in the GLSC database. If not able or willing to meet these conditions, a lab which wishes to have access to the reference genetic standards may still benefit from using these for their own QA/QC and/or comparison with their own data, but will not have access to data in the shared database other than through the scientific literature.

**Table 1:** Microsatellite loci used in this study, showing repeat type, known size range, label used, and source.

<b>Locus</b>	<b>repeat motif</b>	<b>label</b>
<i>Sfo8</i> <sup>a</sup>	dinucleotide	HEX
<i>Sfo12</i> <sup>a</sup>	dinucleotide	FAM
<i>Sfo18</i> <sup>a</sup>	dinucleotide	TET
<i>Sfo23</i> <sup>a</sup>	dinucleotide	FAM
<i>SfoB52</i> <sup>b</sup>	4bp repeat adjacent to 2 bp repeat	FAM
<i>SfoC24</i> <sup>b</sup>	trinucleotide	FAM
<i>SfoC28</i> <sup>b</sup>	trinucleotide	TET
<i>SfoC38</i> <sup>b</sup>	trinucleotide	TET
<i>SfoC79</i> <sup>b</sup>	trinucleotide	HEX
<i>SfoC86</i> <sup>b</sup>	trinucleotide	HEX
<i>SfoC88</i> <sup>b</sup>	trinucleotide	HEX
<i>SfoC113</i> <sup>b</sup>	trinucleotide	FAM
<i>SfoC115</i> <sup>b</sup>	4bp repeat adjacent to 2 bp repeat	FAM
<i>SfoC129</i> <sup>b</sup>	trinucleotide	HEX
<i>SfoC153</i> <sup>b</sup>	trinucleotide	HEX
<i>SfoD75</i> <sup>b</sup>	tetranucleotide	TET
<i>SfoD9.1a</i> <sup>b</sup>	tetranucleotide	HEX
<i>SfoD100</i> <sup>b</sup>	tetranucleotide	HEX
<i>SfoD105</i> <sup>b</sup>	tetranucleotide	FAM

**sources:**

<sup>a</sup>Angers et al. (1995)

<sup>b</sup>Tim King, USGS, Leetown, WV (unpublished)

**Table 2:** Summary of size differences (offsets) in scored allele sizes among labs, compared to data from the GLSC lab. \* Loci *SfoD9.1* and *SfoD105* were discarded after preliminary data comparison among labs.

locus		OMNR (Wilson)	UMinn (Miller)	UWisc (Sloss)
<i>Sfo8</i>		4	5	
<i>Sfo12</i>		0		
<i>Sfo18</i>		2	4	0
<i>Sfo23</i>		-3	-2	3
<i>SfoB52</i>		5		4
<i>SfoC24</i>		2		4
<i>SfoC28</i>		2		3
<i>SfoC38</i>		3		3
<i>SfoC79</i>		0		3
<i>SfoC86</i>		1	2	3
<i>SfoC88</i>		2	2	2
<i>SfoC113</i>		2	3	3
<i>SfoC115</i>		3	3	2
<i>SfoC129</i>		5		4
<i>SfoC153</i>		3		4
<del><i>SfoD9.1*</i></del>				
<i>SfoD75</i>		1	2	1
<i>SfoD100</i>		2	4	2
<del><i>SfoD105*</i></del>				

## LIST OF FIGURES

**Figure 1:** Overview of database model (Stott et al., in prep), showing interface, individual components and overall structure.

**Figure 2:** Brook trout tissue samples and DNA shared among labs to foster data standardization and collaborative efforts. LSC = Leetown Science Center (T. King, USGS); GLSC = Great Lakes Science Center (W. Stott, USGS); UMN = University of Minnesota (L. Miller); OMNR = Ontario Ministry of Natural Resources (C. Wilson); UWI = University of Wisconsin (B. Sloss).

**Figure 3:** Example electropherogram of multilocus microsatellite DNA profile, showing allele phenotypes for di-, tri-, and tetranucleotide repeat microsatellite loci.

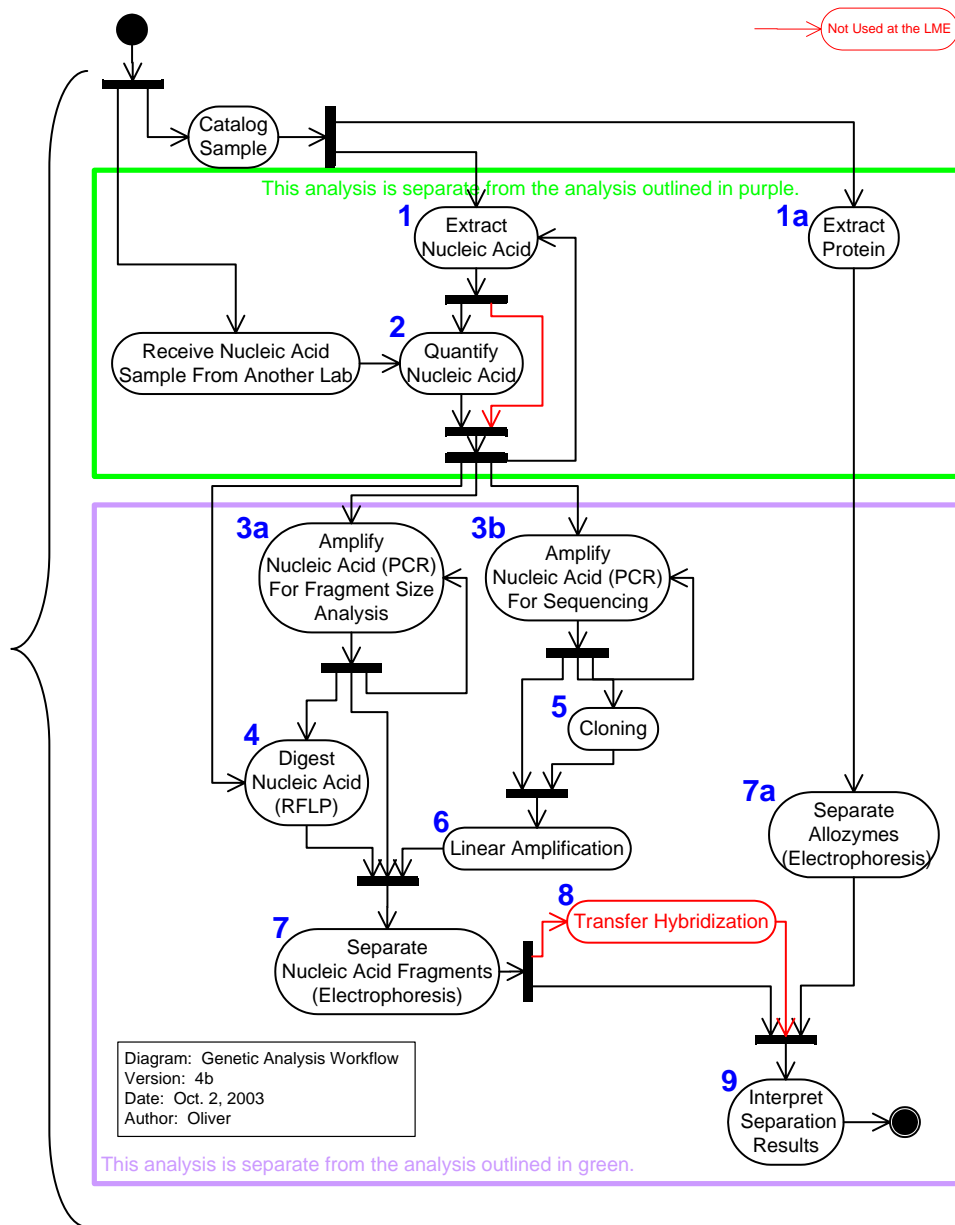
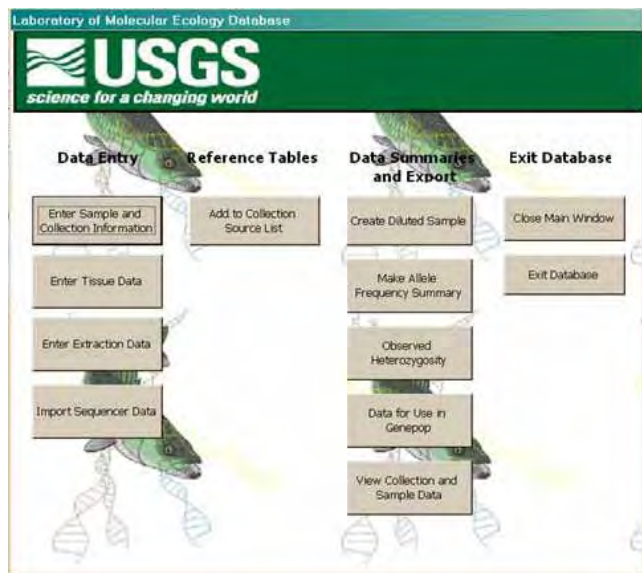
**Figure 4:** Comparison of scored allele sizes for *Sfo23* (Angers et al. 1995), showing estimated allele sizes for first and second allele copies among participating labs.

**Figure 5:** Comparison of overall data congruence based on mean single-locus homozygosity and heterozygosity estimates among labs using raw (uncorrected) data.












**Figure 6:** Allele size comparisons among labs for each microsatellite locus. Data show individual allele size estimates rather than diploid genotypes, and are sorted (rank order by size, smallest to largest allele) using GLSC scores as the primary reference series. Gaps indicate missing data for particular alleles or individuals from a participating lab. Values on horizontal axes represent rank order only.

**Figure 7:** Bivariate plot of scoring consistency among labs, summarizing data from Figure 6. Data points indicate congruence among labs in terms of mean number of alleles detected and percent agreement for allele sizing after applying lab- and locus-specific correction factors. Loci from Angers et al (1995) are indicated by blue diamonds. Error bars show the ranges (maximum-minimum) for each locus and variable.

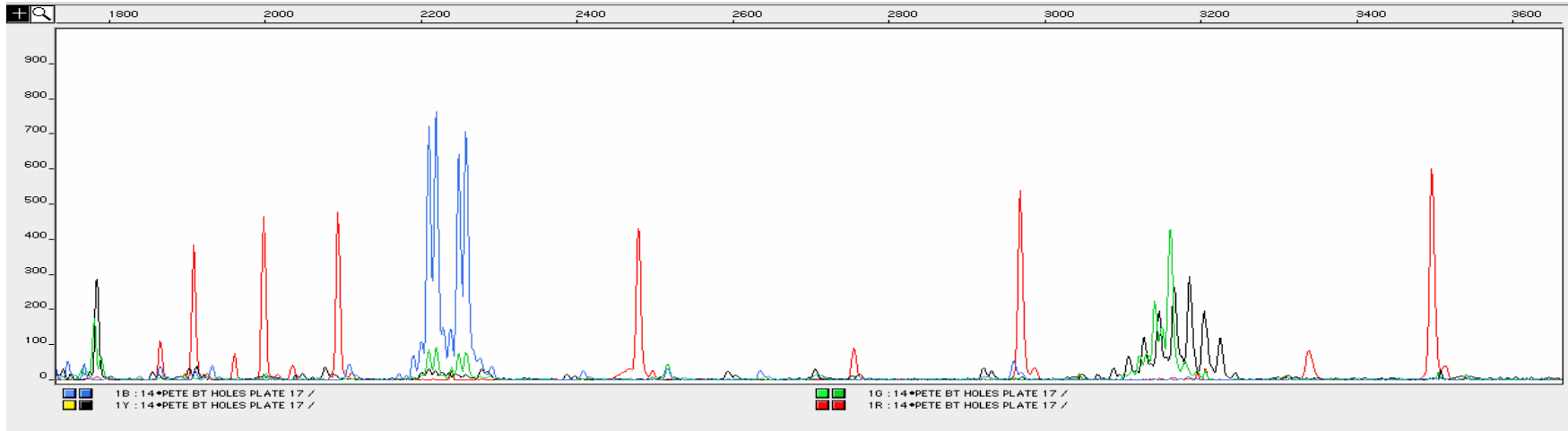
**Figure 8:** Minimum-spanning tree of genetic relationships among Lake Superior brook trout populations, using standardized GLSC and OMNR microsatellite data.



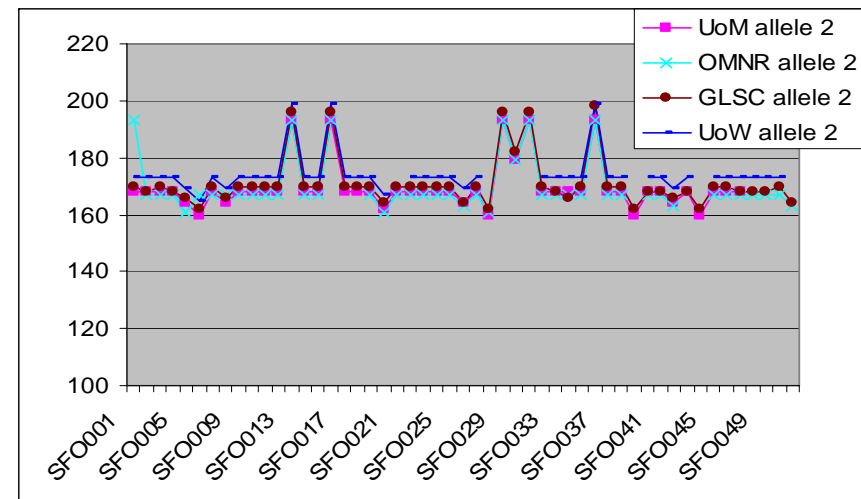
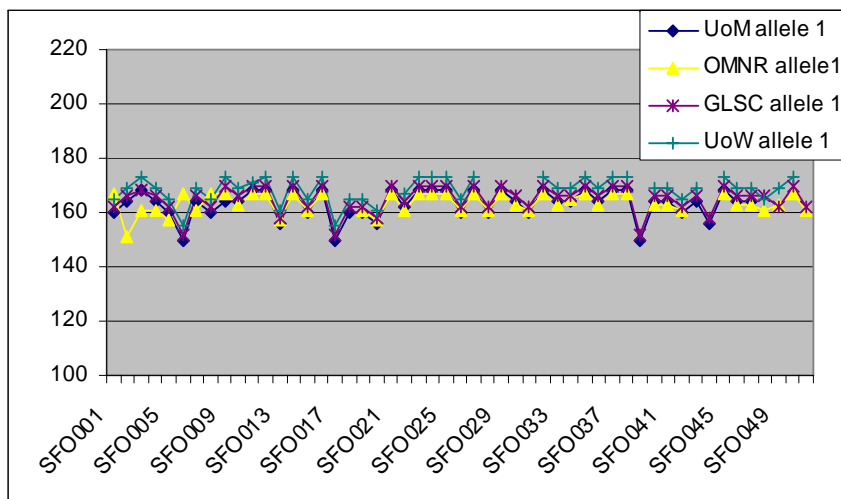
**Figure 1:** Overview of database model (Stott et al., in prep), showing interface, individual components and overall structure.

	LSC	GLSC	UMN	OMNR	UWI
Nipigon broodstock					
Isle Royale, MI					
Grt. Smokey, TN					
Algonquin, ON					

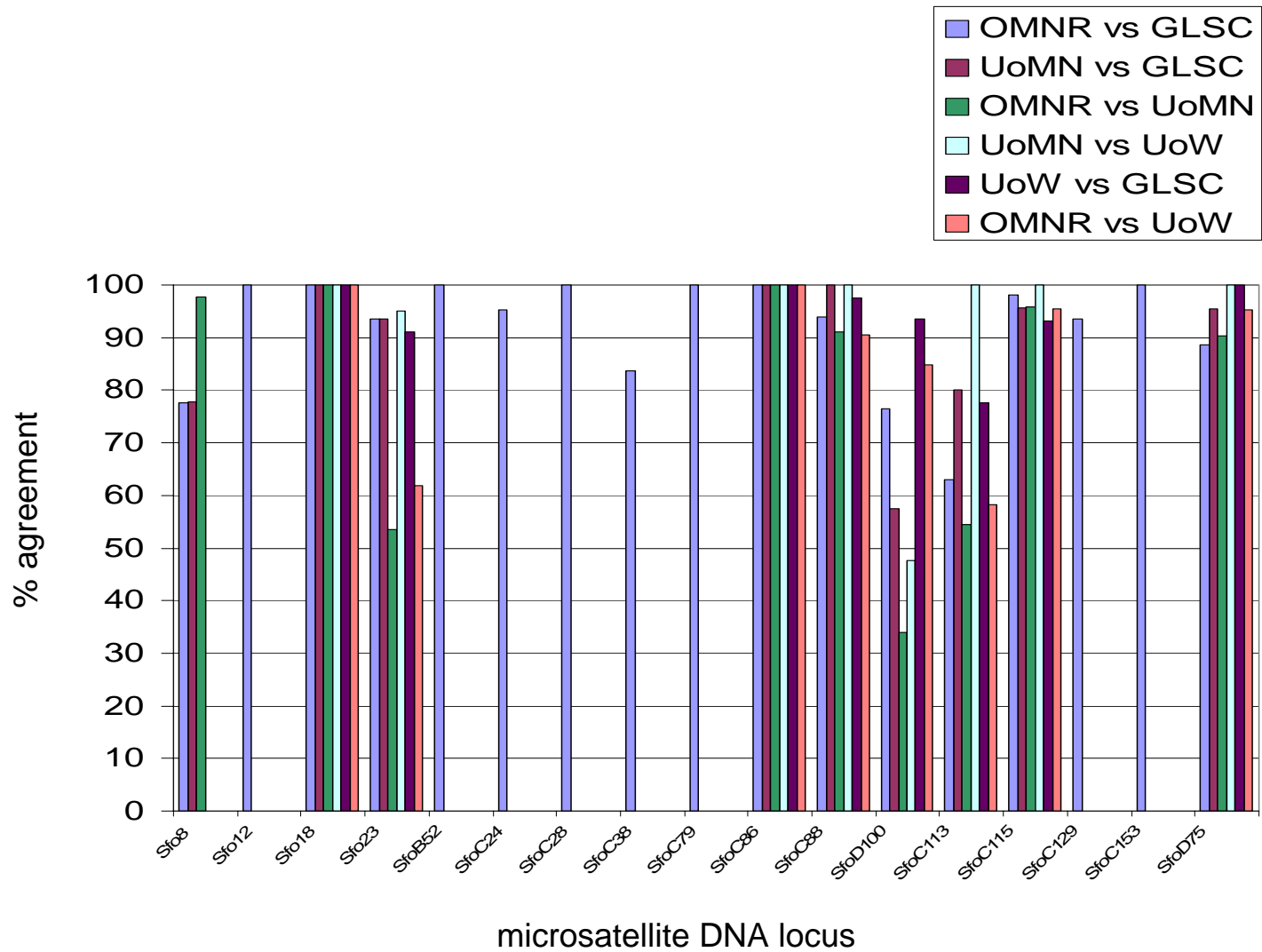
**Figure 2:** Brook trout tissue samples and DNA shared among labs to foster data standardization and collaborative efforts. LSC = Leetown Science Center (T. King, USGS); GLSC = Great Lakes Science Center (W. Stott, USGS); UMN = University of Minnesota (L. Miller); OMNR = Ontario Ministry of Natural Resources (C. Wilson); UWI = University of Wisconsin (B. Sloss).



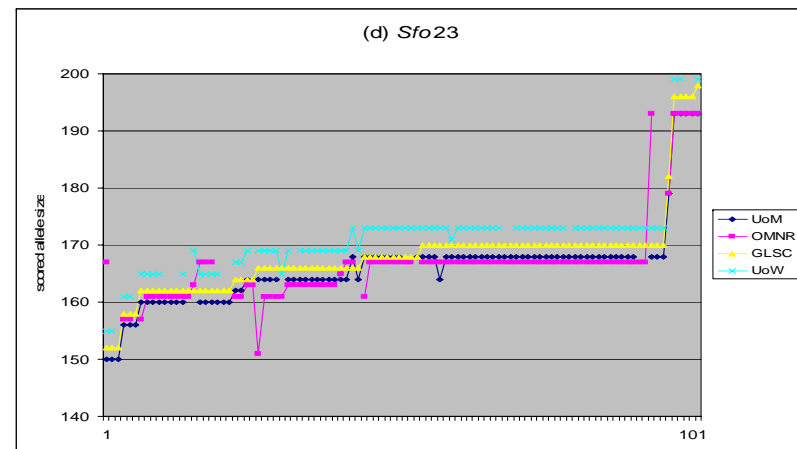
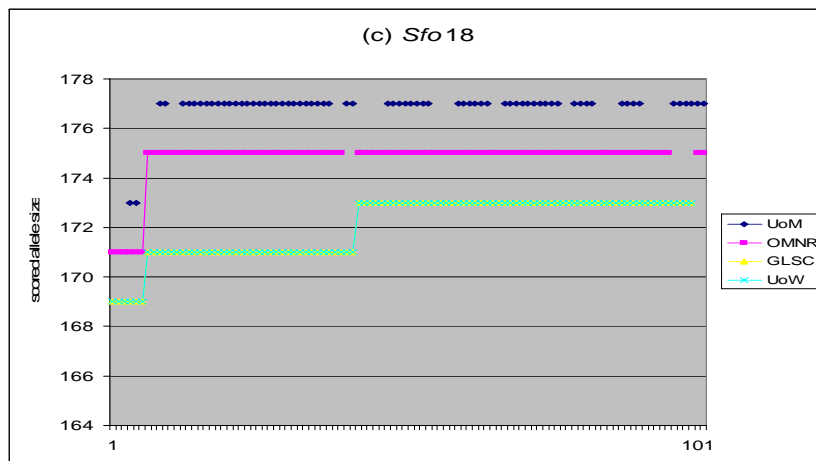
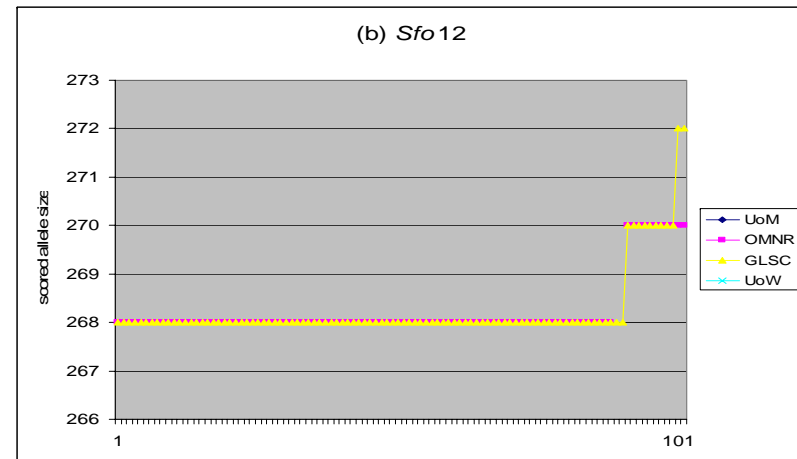
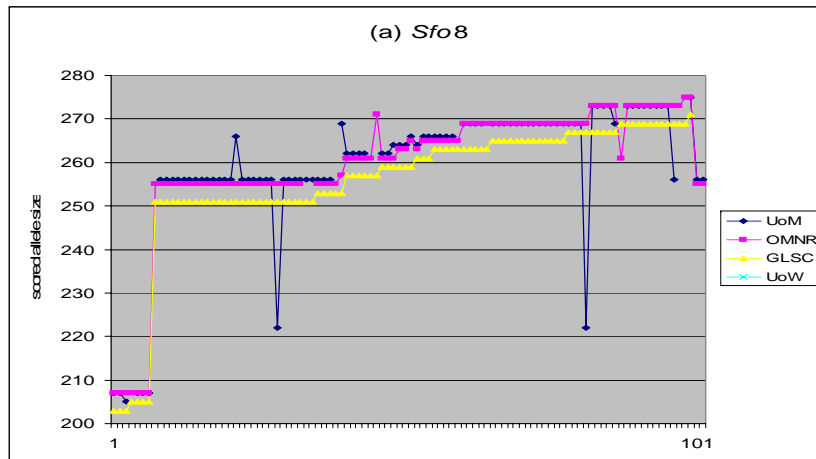
**Figure 3:** Example electropherogram of multilocus microsatellite DNA profile, showing allele phenotypes for di-, tri-, and tetranucleotide repeat microsatellite loci. Red peaks are from an internal size standard (ROX 350®, Applied Biosystems Inc.)



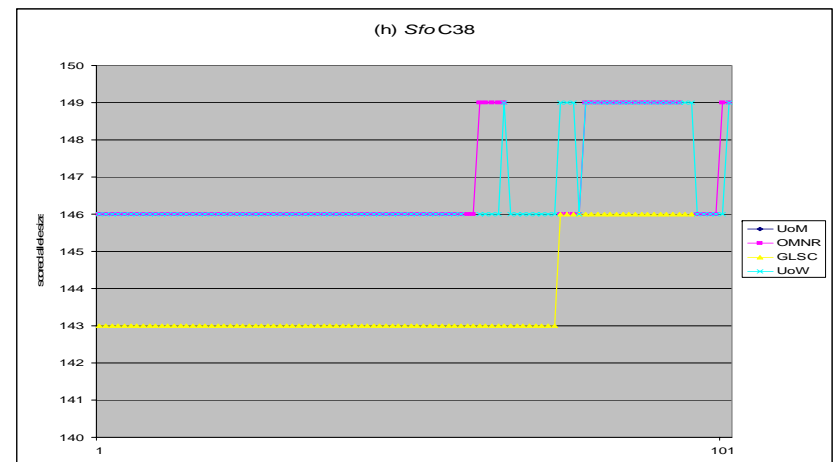
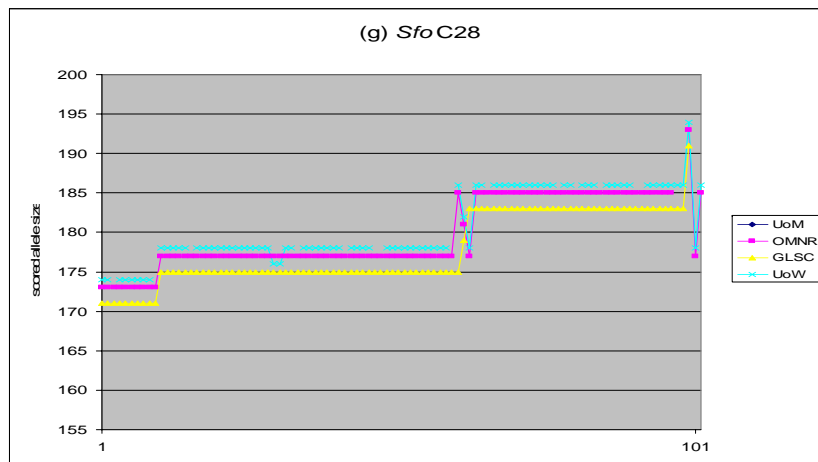
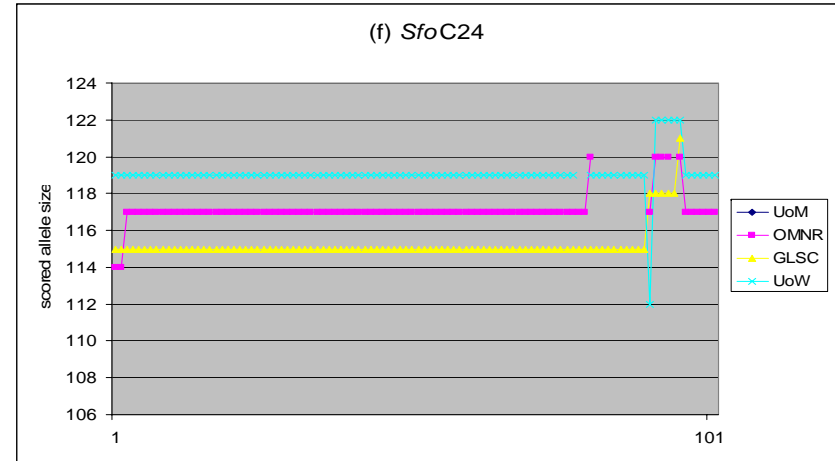
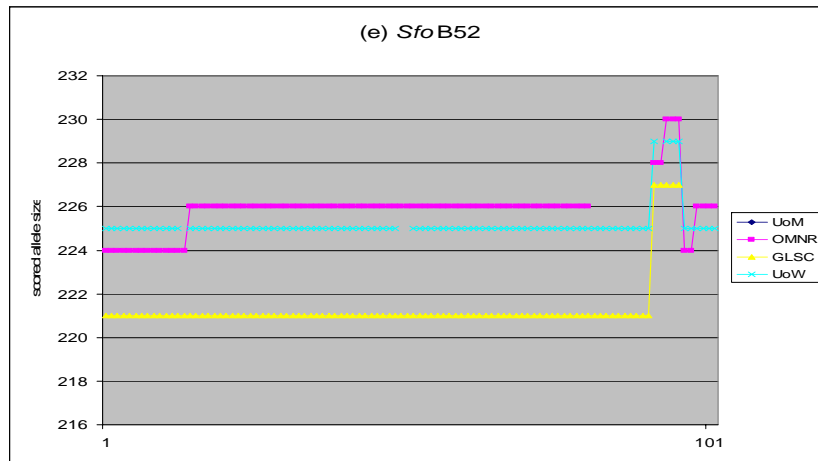
**Figure 4:** Comparison of scored allele sizes for *Sfo23* (Angers et al. 1995), showing estimated allele sizes for first and second allele copies among participating labs. Numbers along horizontal axis represent individual brook trout from the 1999 OMNR Nipigon broodstock cohort.



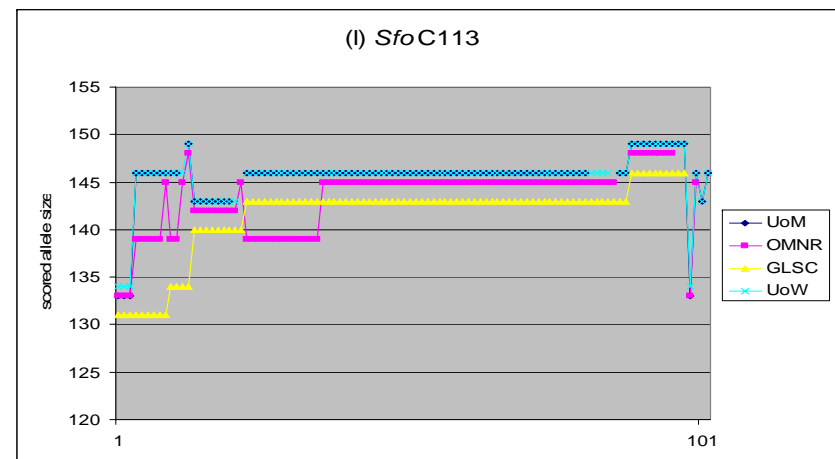
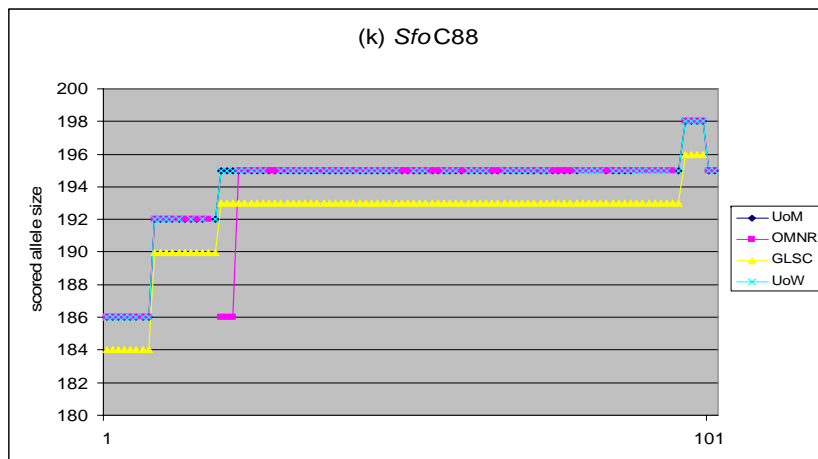
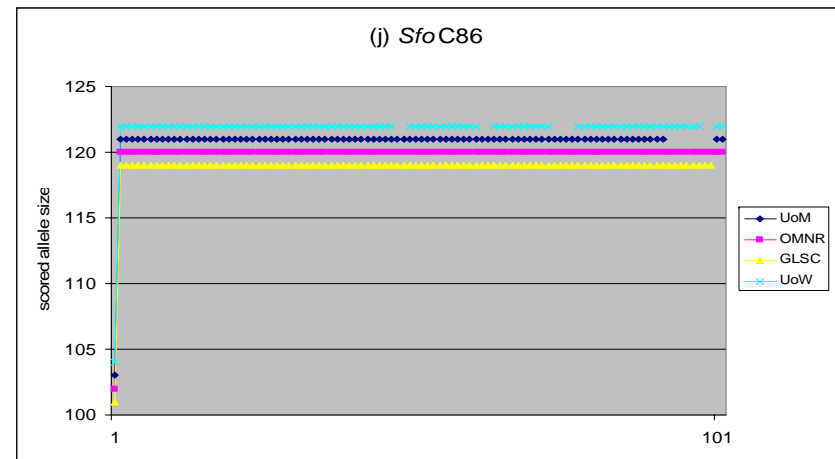
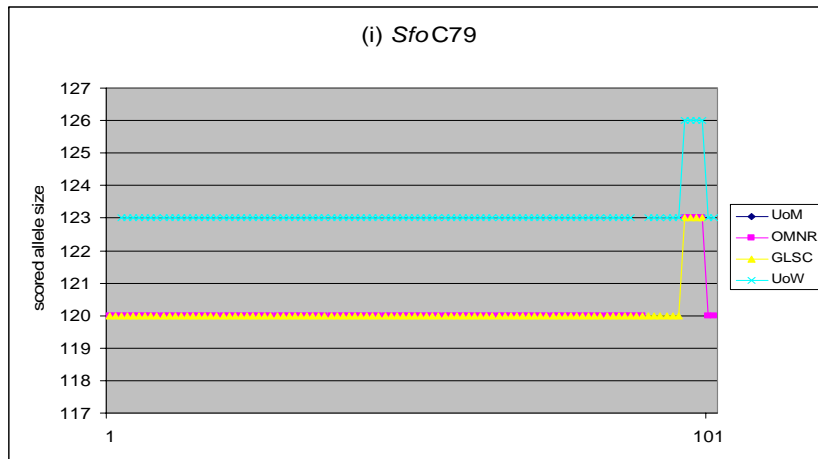
**Figure 5:** Comparison of overall data congruence (percent agreement) based on mean single-locus homozygosity and heterozygosity estimates among labs using raw (uncorrected) data.



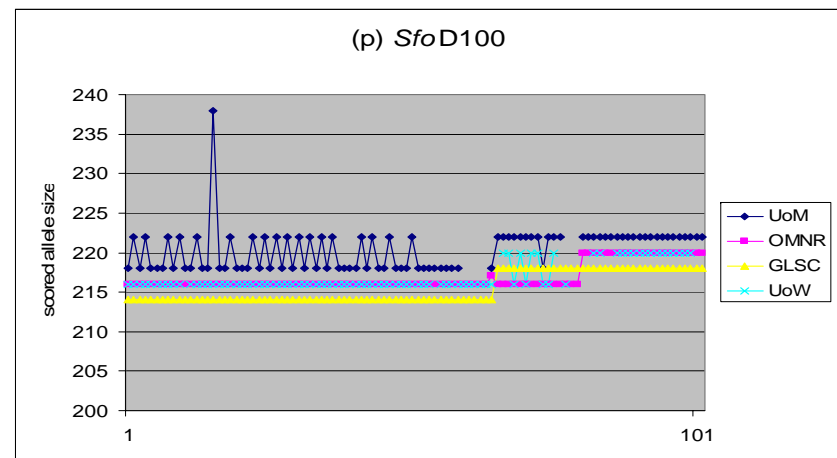
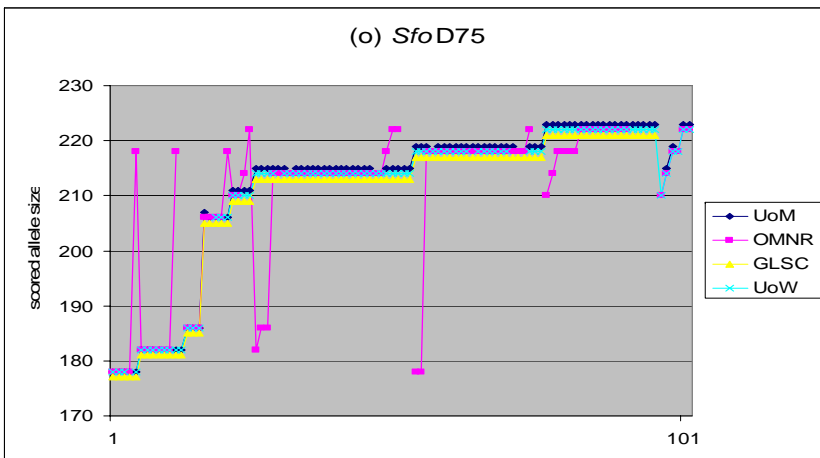
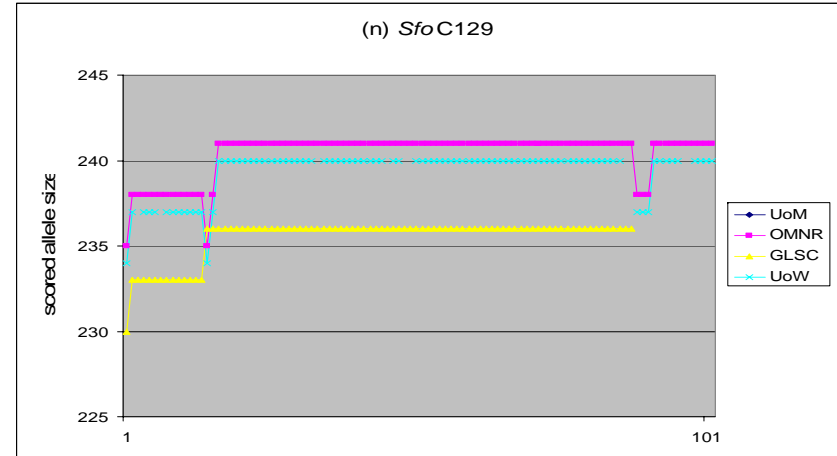
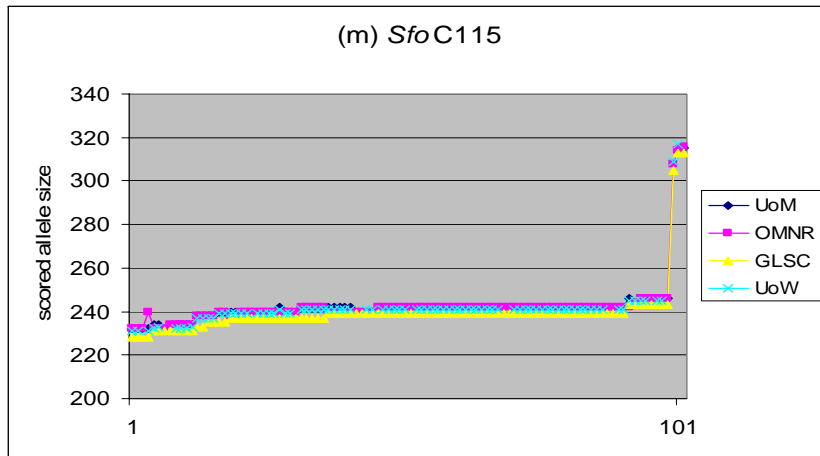
**Figure 6:** Allele size comparisons among labs for each microsatellite locus. Data show individual allele size estimates rather than diploid genotypes, and are sorted (rank order by size, smallest to largest allele) using GLSC scores as the primary reference series. Gaps indicate missing data for particular alleles or individuals from a participating lab. Values on horizontal axes represent rank order only. Microsatellite loci: (a) *Sfo8*; (b) *Sfo12*; (c) *Sfo18*; (d) *Sfo23* (loci from Angers et al (1995)).



**Figure 6:** Allele size comparisons among labs for each microsatellite locus (continued). Data show individual allele size estimates rather than diploid genotypes, and are sorted (rank order by size, smallest to largest allele) using GLSC scores as the primary reference series. Gaps indicate missing data for particular alleles or individuals from a participating lab. Values on horizontal axes represent rank order only. Microsatellite loci: (e) *SfoB52*; (f) *SfoC24*; (g) *SfoC28*; (h) *SfoC38* (loci from T. King, USGS, unpubl.data).

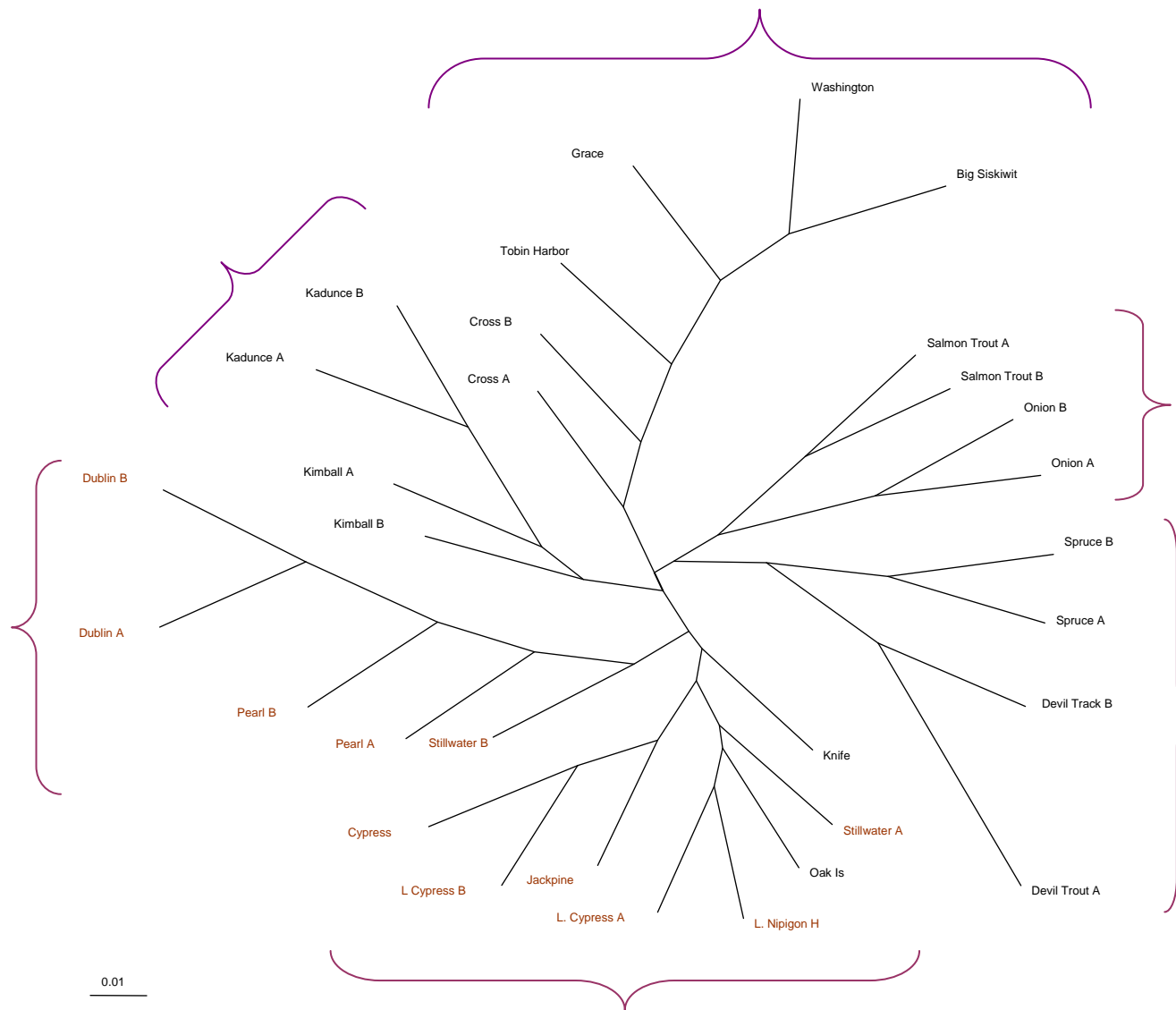


**Figure 6:** Allele size comparisons among labs for each microsatellite locus (continued). Data show individual allele size estimates rather than diploid genotypes, and are sorted (rank order by size, smallest to largest allele) using GLSC scores as the primary reference series. Gaps indicate missing data for particular alleles or individuals from a participating lab. Values on horizontal axes represent rank order only. Microsatellite loci: (i) *SfoC79*; (j) *SfoC86*; (k) *SfoC88*; (l) *SfoC113* (loci from T. King, USGS, unpubl.data).



**Figure 6:** Allele size comparisons among labs for each microsatellite locus (continued). Data show individual allele size estimates rather than diploid genotypes, and are sorted (rank order by size, smallest to largest allele) using GLSC scores as the primary reference series. Gaps indicate missing data for particular alleles or individuals from a participating lab. Values on horizontal axes represent rank order only. Microsatellite loci: (m) *SfoC115*; (n) *SfoC129*; (o) *SfoD75*; (p) *SfoD100* (loci from T. King, USGS, unpubl.data).





**Figure 8:** Minimum-spanning tree of genetic relationships among Lake Superior brook trout populations, using standardized microsatellite data from GLSC and OMNR datasets (W. Stott, USGS unpubl. data; D’Amelio and Wilson, in review).