Stock discrimination of Lake Erie walleye:
A mixed stock analysis contrasting genetic techniques

Timothy B. Johnson[1], Brian Dixon[2], Carol Stepien[3, 4], and Chris Wilson[5]

1. Ontario Ministry of Natural Resources, Lake Erie Fisheries Station, 320 Milo Road, Wheatley, Ontario, N0P 2P0
2. Department of Biology, University of Waterloo, 200 University Ave W, Waterloo, Ontario, N2L 3G1
3. Cleveland State University, Great Lakes Genetics Lab, Center for Environmental Science, Technology and Policy, Cleveland, OH 44115
4. current address: University of Toledo, Great Lakes Genetics Lab, Lake Erie Center, 6200 Bayshore Rd., Oregon, OH 43618
5. Ontario Ministry of Natural Resources, Trent University, 1600 West Bank Drive, Peterborough, Ontario, K9J 8N8

Abstract - Resource managers face continual challenge trying to balance demand with production of fishery resources. Fish populations consist of one to many stocks – discrete spawning units that may exhibit differential survival, growth, and production in response to environmental gradients and exploitation. We designed a blind experiment where 200 walleye *Sander vitreus* of known origin were provided to three different genetics laboratories each charged with identifying the natal origin of each fish (a mixed stock analysis). Three genetic techniques mitochondrial DNA (mtDNA), nuclear microsatellite DNA allelic length frequencies (microsatellites), and nuclear major histocompatibility (MH) gene sequencing were applied to each individual. No single technique was able to correctly assign more than 29% of the individual fish to its natal stock (interlab mean 16.9%). Classification success rose to 53.3% (range 20.7 to 87.9%) when fish were assigned to regions (separate basins within Lake Erie). When techniques were combined, classification success increased marginally. We provide a series of recommendations to both scientists and managers – ensuring that the expectations of each are aligned with technical realities of this rapidly emerging science.

Introduction

Fish stocks throughout the world are in a state of decline (Roughgarden and Smith 1996; Hutchings 2000; Post et al. 2002), yet fish remain the single largest source of protein to the world (Berrill 1997). In North America, the importance of fish as a critical protein source is less, but demand to access these resources by recreational and commercial interests is tremendous. Fishery managers are challenged to provide maximum opportunity to exploit these resources, while ensuring their sustainability. Effective management of fish populations requires an understanding of their structure and behaviour, as well as external factors that influence their survival and production. External factors include everything from climatic events and loss of critical habitat to predator-prey and fishery dynamics. Each of these external factors exhibits spatial variation in its magnitude and relative influence on survival and production. Knowledge of fish populations at smaller spatial scales (i.e., stocks) may therefore improve our ability to effectively manage populations by allowing mitigative or corrective actions to be taken at scales appropriate with the source problem.

Walleye (*Sander vitreus*) is the single most important species to the sport and commercial fisheries of Lake Erie. Since 1950, total annual harvest has ranged from 39,000 kg to 7 million kg, showing dramatic declines in the late 1950s associated with habitat loss and overfishing and an equally impressive recovery in the 1970s following corrective management actions (Hatch et al. 1987; Knight 1997). Since 1990, walleye abundance has again been declining in Lake Erie, and resource managers and scientists are challenged to explain the mechanism(s) behind the decline. Considerable research is investigating the mechanistic effects of nutrient abatement, invasive organisms, altered food web structure, habitat loss, and recruitment variation. Gradually, this research has evolved from whole lake to regional (basin) and now local (stock) scale effects on the lakewide population dynamic (Ryan et al. 1999, Walleye Task Group 2004).

There is considerable evidence that the walleye population of Lake Erie consists of multiple stocks (Regier et al. 1969; Wolfert and vanMeter 1978; Goodyear 1982; Roseman et al. 1996; Mion et al. 1998). Although adult walleye move throughout the lake during much of the year, they separate into discrete spawning stocks during the late winter and early spring in order to spawn (Todd and Haas 1993). Walleye exhibit high spawning site fidelity, returning to the same spawning grounds year after year (Olson and Scidmore 1962; R.C. Haas, Michigan Department of Natural Resources, Mt. Clemens, MI, 48045, unpublished tag-recapture data). Past stock discrimination of Lake Erie walleye relied on a number of techniques. Todd and Haas (1993) were unable to discriminate among spawning stocks in the west basin using electrophoresis. Mitochondrial DNA (mtDNA) analyses have generated mixed results: Billington and Herbert (1988) identified two different mtDNA phylogenetic lineages but were unable to discriminate stocks; Stepien and Faber (1998), McParland et al. (1999), and Gatt et al. (2003) all used mtDNA to distinguish walleye from different geographic areas within Lake Erie, but were unable to separate stocks within a basin; while Merker and Woodruff (1996) found some evidence of discrete breeding groups between the Maumee and Sandusky Rivers in western Lake Erie. Krausse (2002) used major histocompatibility (MH) genes to distinguish Maumee River walleye (west basin) from Cattaraugus Creek (east basin) walleye when investigating the success of a stock transplant experiment. Jarvis et al. (1978) used Fourier analysis of scale shapes to successfully discriminate walleye in eastern Lake Erie, although Riley and Carline (1982) concluded scale shape was a poor discriminator both within and among basins. Hedges (2002) had poor success discriminating walleye stocks when using elliptical Fourier analysis of juvenile walleye opercles. Hedges (2002) and Bartnik (2005) were able to correctly classify 9 discrete spawning stocks of walleye from western and eastern Lake Erie using otolith microchemistry, however, the discriminatory powers were much less when applied to a mixed stock of juvenile walleye.

The objective of this study was to assess the relative power and cost-effectiveness of several genetic techniques (mtDNA, microsatellite DNA and MH genes) in identifying the origins of a set of known-origin spawning walleye in a blind test.

Managing Expectations

One problem when applying population genetic tools for management of wild populations is unrealistic expectations among stakeholders. A common misconception among resource managers and interest groups is that genetic data will precisely identify the source or population membership of pooled samples or individuals. This idea of "genetic barcoding" has been reinforced by popular media, particularly for forensic cases where matching genotypes are used to link or exclude suspects from crime scenes. While genetic membership can usually be resolved among higher taxonomic units (families, genera, and species), this is largely due to the degree of reproductive isolation that occurs at these higher biological levels. For conspecific populations, genetic structure will be complicated by straying or gene flow among populations, resulting in individuals with mixed ancestry and blurred population distinctions. Because genetic data show the ancestry as well as the potential identity of individuals, even small numbers of strays or migrants among separate populations can make identity

difficult to resolve. Furthermore, gene exchange or gene flow (either recent or historical) among separate breeding populations will cause them to appear similar in their genetic characteristics, making it more difficult to recognize distinct genetic units.

The following scenario is a useful analogy for applying genetic markers to identify source populations for individuals. A group of friends at a table each empty their pockets of loose change. Each person counts the number and type of coins that they have, then everyone puts their coins in the centre of the table. If an observer reaches into the pile and pulls out a few coins, which person did those coins come from? Also, how many coins need to be 'sampled' to get an accurate estimate of each person's contribution? This is directly analogous to establishing baseline genetic data for contributing populations, with the pooled coins representing mixed stocks or unknown fish. The ability to correctly identify the source or sources of the coins that were pulled from the pile will depend on the number of coin types, and how many each person put into the pile (stock contributions). For common coins, it's most likely that several people could have been the source; conversely, if only one person at the table was carrying half-dollar coins (50 cent pieces) and one was sampled, the source of that coin can be identified with certainty. If the different people had coins from separate countries, establishing the identity would be fairly simple. Without belabouring the analogy, the more types of coins (alleles) or the greater the variation in their number from different people's pockets, the better the chances of solving the puzzle.

This project attempts to answer a similar puzzle. Any analysis of intraspecific stock structure, or population differentiation and spatial occurrences, is highly dependent on having good baseline data. This requires having a dataset with good spatial (geographic) coverage, sufficiently large sample sizes to obtain representative allele or haplotype frequencies, sufficient variation within the genetic marker system(s) used, and some degree of spatial structure in the resultant data. If any of these conditions are not met, the likelihood of successful application of genetic techniques for management applications is low.

Review of fish stock discrimination techniques

The following section provides a brief description of techniques for fish stock discrimination and directs the reader to primary review articles where more detailed explanation can be found (Table 1). Common molecular biological methods used to identify population genetics and stock structure include allozyme electrophoresis, mitochondrial DNA sequences, and microsatellite DNA. Although the first two methods have lesser potential for stock discrimination than more variable marker systems, these approaches were widely applied to other aspects of fisheries science in the past. We have also briefly summarised several non-genetic based techniques for stock discrimination to better inform resource managers of the array of techniques contained within the broad discipline of genetics.

A. <u>Allozymes</u>

Allozyme/isozyme electrophoresis was the first molecular genetic technique employed to assess genetic structure among populations and species, and is still a widely-used tool. This method is inexpensive, targets metabolically important gene

products, and can readily be applied to a wide array of taxa with little setup cost. Stepien and Kocher (1997) provide an excellent review of this technique. Studies typically examine 10-20 gene loci from frozen eye/brain, muscle, and liver tissues, with more rigourous studies employing more loci. The technique involves the extraction of protein from tissues, their placement in a buffer, and then the running of samples on a gel. These gels are sliced into thin sheets and stained for specific proteins. Results are visualized as bands on gels, with heterozygotes expressing 2 (or more) allelic bands (depending on protein structure) and homozygotes expressing single bands.

Although widely used, allozyme electrophoresis has several drawbacks. To obtain biochemically active biomolecules, tissues must be fresh or flash-frozen and stored at −80°C. Most often, obtaining the necessary tissues involves lethal sampling. Many of the chemicals that are used in staining for target loci are hazardous or toxic. The technique is also less sensitive than more modern techniques, as only genetic differences which result in altered electrophoretic properties of the gene products (allozymes) are detectable. Furthermore, bands (alleles) that have the same electric charge and migrate to the same point in the gel may not be homologous (i.e., may not be comparable). The scoring of gels can be somewhat subjective and bands are difficult to interpret when weak or close together.

Allozyme data have shown geographic structure among walleye populations, but not at a fine enough scale for stock recognition. Of nine allozyme loci assessed by Ward et al. 1989) for walleye across the Great Lakes and northern Manitoba, only one locus showed significant geographic patterning. Todd and Hass (1993) surveyed 25 loci among six spawning walleye populations in Lake Erie and Lake St. Clair, and found significant differences at two loci between lakes, but no significant differences within either lake.

B. Mitochondrial DNA

The mitochondrial (mt) genome has many properties that make it useful for reconstructing recent phylogenetic history, including stock structure, of which the most notable are clonal inheritance and rapid evolution. Fish mitochondrial genomes are haploid and inherited solely from the maternal lineage. They also evolve more quickly than do most nuclear DNA coding regions. One of the advantages of using mitochondrial DNA sequences is that phylogenies of haplotypes can be constructed that trace the geographic and evolutionary patterns of the genotypes over time, potentially discerning several levels of evolutionary history. For example, one can trace the evolutionary and geographic history of walleye across their North American range, and test their relationship to sauger and European pike perch (see Stepien and Faber 1998, Faber and Stepien 1998). A disadvantage is that the entire mitochondrial genome is a single locus (see Stepien and Kocher 1997 for a summary of its properties), and fine-scale population genetic analyses such as stock structure may be best addressed by examining a large number of loci.

Methods involve DNA extraction (using frozen, saline-preserved, or alcohol-preserved tissues), amplification of gene regions using known primers, and DNA sequencing. Sequences are aligned among individuals, and are readily comparable among species and genera. One can thus construct a tree of relationships among the

haplotypes whose pattern can be geographically and evolutionarily analyzed – unlike the microsatellite and allozyme multi-locus methods. Each individual fish has a single mtDNA haplotype (which is the haploid allele). The stock structure data consist of the number and frequency of haplotypes/alleles in the spawning site sample. Analyses test for differences in the number and distribution of haplotypes among samples. In the present study, laboratory A sequenced the mtDNA control region – which is the most rapidly evolving area – and compared the results to analysis of variation at 3 nuclear microsatellite loci.

Early studies of mtDNA variability of walleye in the Great Lakes basin examined fragment patterns generated by digestion of the entire mtDNA genome with restriction enzymes which revealed some large-scale differences and more population differences than did allozymes (Billington and Hebert 1988, Ward et al. 1989, Merker and Woodruff 1996). Later studies which sequenced portions of the mtDNA genome resolved more variability and stock differences (Stepien 1995, Stepien and Faber 1998) than did the RFLP (restriction fragment length polymorphism) analyses (but see Gatt et al. 2000). Examples of mtDNA sequence stock structure studies of walleye showed that individuals are highly variable (Stepien 1995), but the tandemly repeated areas of the mtDNA control region are too variable and may induce evolutionary noise into the analysis (Faber and Stepien 1998). Subsequent studies thus focused on the non-tandemly repeated areas (Stepien and Faber 1998, Stepien et al. 2004, etc.). These studies, with larger sample sizes, found significant differences among some spawning groups in Lake Erie. Notably, older individuals spawning in Cattaraugus Creek showed high genetic diversity and haplotypic divergence reflecting a unique river spawning group (Stepien et al. 2004).

C. Microsatellite DNA

Microsatellites are regions of highly repetitive DNA (e.g., CACACACA…) with unit repeats one to six bases in length that are subject to rapid rates of mutation and high levels of variation. Microsatellites are abundantly distributed throughout the nuclear genome and are highly polymorphic. They are believed to be non-coding and selectively neutral (reviewed in Wright and Bentzen 1994), making them very attractive for studies of population genetic and stock structure. They follow a Mendelian codominant inheritance pattern and allow one to quickly assay a large number of loci. Moreover, advances in molecular genetic methodology allow large numbers of individuals to be assayed for genetic variation, with high accuracy and relatively low cost. Microsatellite DNA provides essentially limitless, highly varied information within a species – ideal for assaying stock structure – because of the potentially large number of loci and individuals that may be surveyed.

Microsatellite data collection involves DNA extraction, PCR (polymerase chain reaction) amplification of known loci using known primers, and sizing of the alleles on a DNA sequencer. Length polymorphisms arising from variation in the number of repeats thus are quantified on the sequencer, in comparison to a size standard. Each individual fish expresses two alleles and may be heterozygous (having 2 different alleles with different lengths) or homozygous (having two alleles of the same length – inherited from each parent). Data are analyzed similar to allozyme studies, for conformance to Hardy-

Weinberg equilibrium expectations and in terms of overall heterozygosity and frequency divergences from other population groups.

Two laboratories involved in this study applied microsatellites, and incorporated both multi-plexed (PCR amplified more than one locus at a time) and pool-plexed (used more than one PCR in a well on a sequencing plate) PCR analysis, which decreases time and expense for data collection. A large number of microsatellite loci have been identified for walleye (Borer et al. 1999, Wirth et al. 1999), and multi-locus microsatellite genotypes were effectively used to assign stocked walleye back to their source populations (Eldridge et al. 2002). Microsatellites provide a means for distinguishing not only among populations but also individuals, allowing assignment tests as well as tracing offspring to parents. The advantage of using microsatellites is increased accuracy and detection of variation as compared to allozyme analysis, due to the capability of assaying large numbers of individuals and loci (see Stepien and Kocher 1997, Smouse and Chevillon 1998, Waser and Strobeck 1998). Disadvantages of microsatellites for genetic analyses include the inability to directly compare levels of diversity across species, and their lack of clear inheritance patterns. Furthermore, their high mutation rate makes them ineffective over broad spatial scales (e.g. range-wide variation). For the purposes of assessing stock structure within the Great Lakes, however, these limitations are not relevant.

D. Major Histocompatibility (MH) genes

Major histocompatibility genes are the most polymorphic known nuclear-encoded genes (Stet et al. 2003). For example, in humans approximately 5000 individuals have been examined and some MH genes have over 500 known alleles (Janeway et al. 2004). MH genes encode proteins that are essential for immune responses to pathogens. These alleles are thought to diversify rapidly in populations due to the extreme selection pressure provided by local pathogens (i.e., individuals with inappropriate alleles die and their alleles are removed from subsequent generations rapidly). Therefore distinct populations of animals should have different sets of MH alleles. The most direct method of assessing MH alleles in different populations is by extracting DNA and sequencing the genes for comparison between individuals or members of different populations. This method has been used to discriminate populations of humans (Parham et al., 1997), geckos (Radtkey et al. 1996) and fish (Dixon et al. 1996, Miller and Withler 1997, Miller et al. 1997). Walleye have one copy of the MH class II alpha gene, thus each individual has two alleles, but they have at least 3 copies of the class II beta gene or up to 6 alleles per individual (Dixon et al. unpublished). The disadvantage of MH gene analyses is the time and expense (about $250 per fish for complete analysis of all alleles present for one gene with 2 alleles). More cost effective methods of discriminating the MH allele complement of a fish are available, such as Single Strand Conformation Polymorphism (SSCP) or Denaturing Gradient Gel Electrophoresis (DGGE), but these come at the cost of less precise determination of sequence differences and should be used only when a solid database of allele sequences is available.

E. Truss analysis

A series of established measurements between fixed reference points on the fish are recorded and analysed in a truss analysis (Strauss and Bookstein 1982). While measurements can be made directly on the fish, a digital image (lateral aspect) not only provides a permanent archive, but enables the fish to be returned to the lake or river unharmed. Interannual variation in body shape (due to differences in prey and growing conditions) can be high and negate comparison across years. Individual variation in body shape and growth may obscure stock specific differences (i.e. variation within a stock may be as great as variation between stocks). No specialised equipment is needed to perform a truss analysis.

F. Shape analysis of hard structures

Any number of hard structures (scales, otoliths, opercles) can be used to perform shape and Fourier analysis (Campana and Casselman 1993, Burnham-Curtis and Smith 1994). Many of these structures are routinely collected by fishery agencies (historic archives may exist). Digital images of structures can be archived, and structures can then be used for other purposes (aging, microchemistry, etc.). Depending on the structure, the sampling can be non-lethal (scales) or lethal (otoliths and opercles). An image analysis system capable of performing the shape analysis is required. Many software packages (Optimus, ImagePro, etc.) have built in routines or macros to assist with the analysis.

G. Otolith Microchemistry

Campana (1999) and Campana and Thorrold (2001) provide excellent reviews of the application of otolith microchemistry to fish stock discrimination. Potential limitations of this technique are the lethal nature of otolith removal (i.e. not well suited for stocks that are imperilled), a limited (but growing) number of facilities that possess the specialised equipment (laser or micro-milling machines attached to ICP-MS) and expertise needed, and the cost (equipment costs and labour intensive). Microelemental concentrations (stock signatures) can vary between years.

Table1 provides a brief synopsis of each technique, identifying the potential for stock discrimination, the tissue needed, the primary strengths and weaknesses, and a crude estimate of cost.

Table 1.- Comparison of common techniques used to discriminate fish stocks.

Technique	Stock Resolution	Tissue needed	Primary Strength	Primary Weakness	Unit Cost (in 2004 US dollars)	Primary Reference
allozymes	suited only for coarsest levels of population subdivision	eye/ brain, liver or muscle	Large number of loci available; low set-up costs	not variable enough for stock discrimination	$15 per individual for 50 loci	Stepien and Kocher 1997; Hills et al. 1996
mtDNA	low to moderate due to single locus; variation limited to DNA mutation rate	DNA sample	Higher-level geographic problems, phylogenetic comparisons	Most walleye in Lake Erie belong to common haplotypes due to shared postglacial ancestry	$25-50 per individual for about 1500 base pairs	Stepien and Faber 1998
Microsatellites	high due to large number of available loci	DNA sample	highly variable; lots of alleles and loci	Population structure difficult to interpret if stocks interbreed	$25 for 15 loci once data base is built	Sunnucks 2000
MH genes	very high	DNA sample	provides very high resolution	very expensive; may be too variable (detect individuals rather than stocks)	$250 per individual per gene	Dixon et al. 1996; Miller and Withler 1997; Miller et al. 1997
Truss analysis	variable	digital image (lateral aspect)	no specialised equipment or costs	within stock variability is high	negligible	Strauss and Bookstein 1982
Shape analysis	variable	otolith, scale, opercle	structures are routinely collected by agencies	within stock variability is high (but less than with truss)	initial equipment purchase only	Campana and Casselman 1993
Otolith microchemistry	high	otolith	time recording properties of otoliths greatly enhance this technique beyond stocks (aging, movement, etc.)	limited facilities; lethal sampling	$30 per individual	Campana and Thorrold 2001

Methods

Experimental Design

In March 2002, 100 spawning walleye were sampled at each of three locations in Lake Erie: the Maumee River (Ohio), Hen Island shoal (Ontario), and van Buren reef (New York) (Figure 1). The identity of the sites (stocks) was known only to the project coordinator. These three sites were chosen (from the 12 or more known discrete walleye spawning locations in Lake Erie) because 1) they represented prominent walleye stocks in Lake Erie, 2) limited tag-recapture data (Bob Haas, Michigan DNR, unpublished data) suggested discreteness or spawning fidelity at these sites, 3) the sites were sufficiently geographically separated to minimise potential straying of fish (which would confound the assumed 'discreteness' of the stocks), and 4) they represented both river and shoal stocks both of which are assumed important contributors to overall production but may have different genetic lineages associated with these separate spawning behaviours. We limited the potential sites to three to

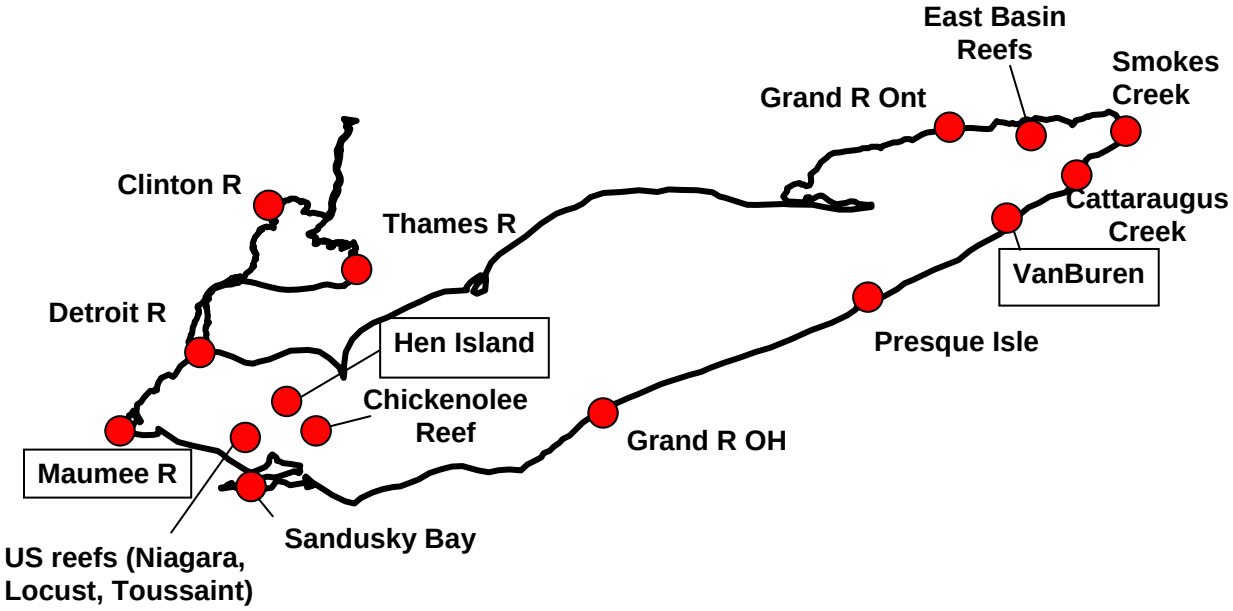

Figure 1 – Probable spawning stocks of walleye in Lake Erie. Stocks used in this analysis are outlined.

ensure sufficient statistical power to assess the accuracy of the assignments while acknowledging the intra-stock (individual sample) variability associated with the assignment. Only sexually mature (ripe) fish were used, but no additional effort was made to control for size or sex of fish. From each fish, the distal 1 cm of the pectoral fin was removed and placed in 85% ethanol. Fin sections were used as they represented non-lethal tissues (as opposed to vital organs or otoliths). All samples from a given collection site were placed in a single, well labelled jar and forwarded to the Ontario Ministry of Natural Resources (OMNR) office in Wheatley.

Once samples were obtained from all three collection sites, 200 samples were randomly drawn across the three stocks. Individual vials were assigned unique sample

ID numbers, and one fin section from the appropriate stock was placed in each vial, again containing 85% ethanol. Again, only the project coordinator knew the stock origin of the fish; the genetic laboratories knew only the samples represented walleye from Lake Erie (i.e., a blind, mixed-stock analysis). The samples were then forwarded to the OMNR Fish Genetics Laboratory in Peterborough for DNA extraction. We chose a single lab to perform the extractions to minimise variation associated with this procedure that would confound the intended purpose of comparing genetic stock discrimination techniques among laboratories.

Three replicate samples of DNA were extracted from each fish (see below for detailed methods of extraction), maintaining the original sample ID number. One replicate sample of the DNA was then sent to each of the three participating laboratories. Each laboratory was asked to handle the sample as they would any other sample sent to them (i.e. no additional analyses, replication, or other handling that would make these sample assignments more precise than a 'normal' sample). Our experiment was designed to mimic a mixed stock analysis, where a sample of fish of unknown origin would be analysed to determine the natal stock for each individual fish. Therefore, the specific instruction to each laboratory was to provide the single most likely stock identity as well as an indication of the confidence in that assignment. If the fish could not be assigned to a single stock (stock signature showed characteristics of multiple stocks), then the laboratory should provide the smallest spatial unit that could describe the identity of that fish (i.e. west basin reef as opposed to a west basin river, or west basin versus eastern basin, etc.). Once all fish had been assigned to probable stocks, the result was forwarded to the project coordinator who compared the assignment with the known stock origin. The identity of the stocks, and of the individual fish, was not revealed to any laboratory until all genetic analyses were completed.

DNA extraction:

Total DNA was extracted from 40 mg subsamples from the preserved finclip tissue per fish, using a standard organic extraction protocol (Sambrook et al. 1989, Bardakci and Skibinski 1994). Minced tissue was broken down in a lysis buffer of 500 μL STE with the addition of 30 μL SDS (10% w/v) and 10 μL of proteinase K (2 mg/mL). Proteins were removed using sequential extraction with 500 μL washes of phenol and chloroform. DNA was precipitated using 1 mL of 95% ethanol, rinsed with 500 μL 70% ethanol, and resuspended in 200 μL of TE elution buffer. Quality and yield of extracted DNA was confirmed by horizontal gel electrophoresis alongside molecular standards, and any samples with insufficient quantity or quality were replaced with other unknowns provided by T. Johnson. Eluted DNA was separated into triplicate 60 μL aliquots, and identical sets were sent to each of the participating labs.

DNA Analyses:

Laboratory A: mitochondrial and microsatellite DNA

The entire mitochondrial DNA control region (also called the D-loop), between the tRNA proline to tRNA phenyalanine regions, was amplified and sequenced using conserved primers and following methods previously described (Stepien 1995, Stepien and Faber 1998, and Stepien et al. 2004). Sequencing was performed using a

Beckman-Coulter CEQ 8000 autosequencer. Sequences were aligned against the most common haplotypes of walleye in Lake Erie (Stepien and Faber 1998) using the Beckman-Coulter CEQ software. We omitted the sequential repeat and imperfect repeat regions at the 5' end of the control region, which are described in detail in Faber and Stepien 1998.

A total of 19 mtDNA control region haplotypes (excluding the 5' sequential 11 bp tandem repeat; see Faber and Stepien 1998) were identified for walleye sampled from Lake Erie in the Great Lakes Genetics Laboratory baseline data base, which included a total of 272 individuals, including 64 from the Maumee River, OH; 47 from the Sandusky River, OH; 20 from Sandusky Bay, OH; 15 from U.S. west basin reefs, OH: 32 from the central basin rivers (Grand and Chagrin), OH; 55 from Cattaraugus Creek, NY, 50 from Van Buren Bay, and 39 from the Grand River, Ont.

In addition, variation was assayed at 3 nuclear DNA microsatellite loci (*Svi*33, *Svi*4, and *Svi*18) from the same baseline individuals, using primers described by Borer et al. (1999). Some of the mtDNA data were incomplete for the microsatellites, and their inclusion reduced our baseline to 248 walleye overall. Microsatellites were amplified using the polymerase chain reaction (PCR) with two flanking primers (one of each pair was dye-labeled). PCR reactions were optimized for resolution using a MJ Research gradient temperature thermalcycler. The PCR products were then run on a Beckman-Coulter CEQ 8000 capillary autosequencer and fragment lengths were resolved and scored by computer in relation to the dye standard. All print-outs were manually checked and the files were stored. Heterozygotes were identified as having two peaks of different lengths and homozygotes had a single peak of greater amplitude. The laboratory both multi-plexed (amplified two or more microsatellites per PCR reaction) and pool-plexed (ran more than one PCR reaction in a single well on the sequencer), in order to increase output and decrease cost. Several different Beckman dyes were employed, which combined with the fact that many of these microsatellite loci do not overlap in length range of alleles, allowed several loci to be resolved in single PCRs (multi-plexing) and in single wells of the microplate (pool-plexing). We utilized WhichRun version 4.1 (Banks and Eichert 2000 and http://www.bml.ucdavis.edu/whichrun.htm) to test for population assignment of individuals based on multilocus genotype data, combining the mtDNA haplotype with the 3 microsatellite DNA loci data. In the case of the former, the data were entered as homozygous. The WhichRun program assumed that each baseline population had Hardy-Weinberg-Castle genotype frequencies (excepting the mitochondrial data) and that the genetic loci employed were independent. (Hardy-Weinberg equilbria and loci independence were independently tested and verified using the computer program Arlequin 2.2 (Schneider et al. 2003)). The WhichRun program also presumed that the likelihood that an individual sample came from each of the source populations was equal to the H-W-C frequency of its specific genotype at each locus in each respective source population. Likelihood values for each locus were multiplied to give a series of multi-locus likelihood functions for assignment to each of the source populations. The jackknife subroutine in WhichRun was used to provide an empirical means for evaluating baseline data and the chances of correct allocation. Each iteration sampled individuals from the baseline one at a time, recalculating allelic frequencies in the

absence of each individual genotype sampled before determining most likely population origin for that individual.

Laboratory B: MH alleles

DNA samples were used as templates for PCR amplification of the variable part of MH class II alpha genes. PCR fragments were cloned into the plasmid vector PGEMT-easy and transformed into the DH5 α strain of *E coli*. After growth of the culture, plasmids were extracted and used for DNA sequencing using the dideoxy chain termination method on a Visible Genetics long read tower. Sequences were assigned allele numbers and were used to compile a dataset for each of the 200 individuals. This dataset was used for assigning individuals to predefined populations using several methods available in the program Geneclass v2.0 (Piry et al. 2004). An analysis of the number of populations present was attempted using the program Structure (Pritchard et al. 2000).

Laboratory C: 2 separate microsatellite DNA multiplexes

Resuspended DNA was quantified using gel electrophoresis alongside a mass ladder standard (MassRuler®, MBI Fermentas) and standardized to a working concentration 30 ng/μL. Multilocus genotypes were amplified from individual fish using two multiplex polymerase chain reactions (PCR). The first 10 μL PCR reaction co-amplified microsatellite loci *Svi*2, *Svi*4, *Svi*6, *Svi*7, and *Svi*14 using primers developed by Borer et al. (1999), 0.05U *Taq* DNA polymerase, 1.5 mM Mg^{++} and 0.5 μL of 10mM dNTPs. The second multiplex PCR reaction amplified microsatellite loci *Svi* L4, *Svi* L5, *Svi* L6, *Svi* L7, *Svi* L8, and *Svi* L11 (Wirth et al. 1999), using the same reactant amounts. For both multiplex reactions, each locus was labeled with a fluorescent tag to permit simultaneous visualization on an automated sequencer. Primer concentrations ranged from 2.5 pM to 5.0 pM in each reaction, depending on the relative amounts of amplification products.

Thermal cycling conditions consisted of an initial heating step at 95°C step for 11 minutes, followed by 30 cycles of denaturation at 94°C for 1 minute, holding at primer annealing temperatures for 1 minute (54°C for reaction 1; 58°C for reaction 2), and extension or DNA synthesis at 72°C for 1 minute. A final extension step at 60°C for 45 minutes greatly improved the quality of the amplified data.

The PCR products were visualized on an Applied BioSystems 377XL automated DNA sequencer using 36cm 5% Long Ranger polyacrylamide gels and an internal ROX 350 size standard (Applied BioSystems). Product from both PCR multiplex reactions were diluted 1:4 in TE prior to visualization on the sequencer. Data were extracted using GeneScan v3.1.2, scored with GenoTyper v2.5 (Applied BioSystems), and confirmed by manual proofreading. The final data set of individual multilocus genotypes was tabulated and binned for statistical analysis.

The predictive power of recognizable stock structure relative to baseline data (Wilson et al. in prep.) was assessed in several ways. First, although spawning aggregations of fish are usually considered to be single cohesive populations and/or distinct "stocks", this is not a safe assumption. All spawning groups were therefore first tested for conformation to Hardy-Weinberg equilibrium expectations for single panmictic

populations. Relatedness and divergence among putative stocks was assessed using neighbour joining of genetic distances to identify the number of major genetic groups (data not shown). Genic differentiation among spawning runs across Lake Erie was calculated using GenePop 3.4 (Raymond and Rousset 1995). Pairwise F_{ST} estimates among river and shoal-spawning walleye runs were calculated using FSTAT 2.9.3 and tested for significance using 5,000 bootstrap iterations (Goudet 2001). Lastly, spawning fish from baseline populations were tested for population membership using individual assignment tests using GeneClass2 (Piry et al. 2004), using a Bayesian "leave one out" sampling algorithm (Cornuet et al. 1999).

The 200 test samples were then assigned to most-probable source using individual assignment tests implemented in GeneClass2 (Piry et al. 2004), using available data from known spawning runs. Source populations for the set of unknown-origin walleye were estimated using several groupings of the baseline data, as outlined below:

(1) To maximize spatial resolution, all spawning runs with available data were considered as potential sources and included as separate entities or stocks;
(2) Spawning runs were kept as separate sources, but Cattaraugus Creek (NY) and pockets of shoal-spawning fish from eastern Ontario waters were excluded, as both violate expectations for single cohesive genetic groups (unpubl. data);
(3) Walleye from separate geographic locations that were not statistically different were grouped as genetic stocks (e.g. van Buren Bay and Smokes Creek (NY); Maumee River, Sandusky Bay, and Grand River (OH))
(4) Baseline data from spawning runs were grouped into three geographic genetic stocks (Grand River (ON); native New York populations (van Buren Bay and Smokes Creek (NY); and west & central basin spawning runs).

Results from run 1 (stocks) and run 4 (groups) are presented n this analysis.

Results

<u>Discrimination at the level of stock:</u>

Of the 200 spawning walleye represented in the sample, 66 originated from Hen Island (west basin reef), 76 from the Maumee River (west basin river), and 58 from VanBuren reef (eastern basin). For all results, we consider the site of collection to represent a discrete stock. At the level of stock, agreement between labs was low (Table R1), with no more than 9% of the individual fish being assigned the same stock by two labs, and only 1% of the fish being assigned to the same stock by all three labs. When compared to known stock origin, agreement between labs was ≤ 1%, and no fish was correctly assigned by all three labs (Table R2).

Overall correctness of stock assignment varied between laboratories (Table R3) in part due to the number of stocks represented within each laboratory's library. Ten stocks were represented by the three labs, although each individual lab contained only 7 (lab C) or 8 (labs A and B) stocks. Correct stock assignment ranged from 1.3 to 28.9%, with each lab showing a higher classification success for a different stock.

Table R1.- Pair-wise comparison of walleye stock assignment among three laboratories. Values are the % of cases where the same fish was assigned to the same stock by both or all three laboratories.

	Laboratory B	Laboratory C	All
Laboratory A	3	9	
Laboratory B		5	
All			1

Table R2.- Correctness of assignment of pair-wise comparison of walleye stock assignment among three laboratories. Values are % of cases where the same fish was assigned to the same stock by both or all three laboratories.

	Lab A & B	Lab A & C	Lab B & C	All
Hen Island	0	0	0	0
Maumee River	0.5	0	0.5	0
VanBuren Reef	0.5	1	1	0

Of equal interest to correctness is any pattern of misclassification. Lab A assigned a high proportion of fish to Lake St. Clair and Sandusky Bay irrespective of natal origin, and a high proportion of fish to the Grand River (Ont) for fish of west basin origin (Hen Island and Maumee River). Lab B consistently assigned >35% of the fish to the Presque Isle (central basin, PA), while east basin stocks (VanBuren, Smokes Creek, Cattaraugus Creek, and Grand River (Ont)) were consistently assigned lower proportions of fish, irrespective of natal origin. Lab C assigned a high proportion of west basin stock fish (Hen Island and Maumee River origin) to Grand River (OH) and Sandusky Bay, while many of the misclassified VanBuren reef fish were assigned to Smokes Creek, an east basin tributary.

Cattaraugus Creek is an interesting stock, as Maumee River walleye were stocked into Cattaraugus Creek from 1995 to 2000 to rebuild the stock. For fish of known Maumee origin, no lab misclassified many as Cattaraugus fish (which one would assume would have a strong Maumee signature).

Discrimination at regional scales

Due to known uncertainty in genetic stock discreteness (fish straying between sites, sampled fish not representative of the spawning unit, etc.), each laboratory was asked to provide assignments to the finest level of resolution they felt could be confidently discriminated. In other words, existing libraries are not suitable for discriminating at the level of stock. Laboratory A defined to spatial units representing west basin and east basin walleye. Laboratories B and C, further subdivided the east basin into north shore (Grand River (Ont)) and south shore (New York) groups. On the basis of these aggregated stock assignments, agreement between labs increased markedly (Table R4). Agreement for correctly assigned fish was much higher for west basin walleye than east basin (VanBuren reef) walleye (Table R5). Labs A & C agreed

on the correct assignment of 37% of the west basin fish, while all three labs correctly agreed on the assignment of 13% of the west basin fish.

Table R3.- Stock assignment (number of fish) for each laboratory compared against the known stock identity. When the known stock and the assigned stock bear the same name, the assignment was correct; if not the assignment was incorrect. 200 individual walleye were used in the experiment with unequal representation among the three known stocks. n/a indicates that stock was not available in the library.

Known Stock	Assigned Stock	Lab A	Lab B	Lab C
Hen Island (n=66)	Cattaraugus Creek	1	0	n/a
	Grand R Ont	18	6	3
	Grand R OH	7	n/a	16
	Hen Island / western reefs	**14 (21.2%)**	**11 (16.7%)**	**10 (15.2%)**
	Maumee River	2	19	5
	Presque Isle	n/a	25	n/a
	Sandusky Bay	9	1	14
	Lake St. Clair	13	n/a	n/a
	Smokes Creek	n/a	0	4
	Van Buren reef	2	1	14
Maumee River (n=76)	Cattaraugus Creek	3	0	n/a
	Grand R Ont	12	10	3
	Grand R OH	10	n/a	13
	Hen Island / western reefs	15	9	7
	Maumee River	**1 (1.3%)**	**22 (28.9%)**	**13 (17.1%)**
	Presque Isle	n/a	27	n/a
	Sandusky Bay	13	1	17
	Lake St. Clair	17	n/a	n/a
	Smokes Creek	n/a	0	9
	Van Buren reef	5	7	14
Van Buren (n=58)	Cattaraugus Creek	3	0	n/a
	Grand R Ont	5	11	2
	Grand R OH	5	n/a	7
	Hen Island / western reefs	1	5	4
	Maumee River	0	9	6
	Presque Isle	n/a	22	n/a
	Sandusky Bay	24	3	8
	Lake St. Clair	13	n/a	n/a
	Smokes Creek	n/a	0	16
	Van Buren reef	**7 (12.1%)**	**8 (13.8%)**	**15 (25.9%)**

Table R4.- Pair-wise comparison of walleye regional assignment among three laboratories. Values are the % of cases where the same fish was assigned to the same regional group of walleye by both or all three laboratories.

	Laboratory B	Laboratory C	All
Laboratory A	22	53	
Laboratory B		29	
All			14.5

Table R5.- Correctness of assignment of pair-wise comparison of walleye aggregate stock assignment among three laboratories. Values are % of cases where the same fish was assigned to the same regional group of walleye by both or all three laboratories.

	Lab A & B	Lab A & C	Lab B & C	All
west basin (Hen + Maumee)	17.5	37	16	13
VanBuren Reef	2	7.5	4	0.5

Overall correctness of aggregate stock assignment again varied between labs (Table R6). Lab A had very high classification success for west basin walleye (85.2%) with lower success (43.1%) for east basin walleye. Lab B had lower success for all three groups of fish, although classification for east basin fish was much higher (84.5%) if both east basin sub-groups (Grand R (Ont) and New York) were combined. Lab C was much more consistent across the lake with respect to correctness of assignment (55.3 to 69.7%).

Table R6.- Aggregate stock assignment for each laboratory compared against the known stock identity. Values are the number of individuals assigned to each stock. 200 individual walleye were used in the experiment with unequal representation among the three known stocks.

Known stock	Assigned Stock	Lab A	Lab B	Lab C
Hen Island	**west basin**	**58 (87.9%)**	**19 (28.8%)**	**46 (69.7%)**
(west basin)	Grand R Ont	n/a	38	3
(n=66)	New York	8	6	17
Maumee River	**west basin**	**63 (82.9%)**	**22 (28.9%)**	**42 (55.3%)**
(west basin)	Grand R Ont	n/a	44	3
(n=76)	New York	13	10	31
VanBuren Reef	west basin	33	9	19
(New York)	Grand R Ont	n/a	37	3
(n=58)	**New York**	**25 (43.1%)**	**12 (20.7%)**	**36 (62.1%)**

Recognising that the techniques applied in this study are complementary, we had originally intended to carry out composite assignments, first pairing techniques (laboratories) and then pooling all three techniques into a single matrix. These results may be indicative of the relative incremental increase in correctness of stock assignment when multiple techniques are employed. Unfortunately, it was not possible to assess the cumulative power and cost-effectiveness of combining the different genetic marker systems. Differences in stock coverage for baseline data among the different laboratories prevented building a composite baseline dataset or library. Two approaches to build a suitable composite were attempted without success. First, baseline data for each lab were pooled only for source populations ('stocks') that were shared among all three laboratories. The large discrepancies in baseline data sample sizes among labs, particularly for MH variation, caused significant problems, as did populations such as Lake St. Clair and Hen Island for which only one or two labs had baseline information. A second approach was to simulate composite genotypes to build a synthetic dataset. Although this would circumvent weaknesses caused by different samples, this approach introduced substantial bias. We are still pursuing options to merge data types, to better assess their cumulative power, but this will require some sample sharing among labs in order to have at least some baseline data from each potential source within the different lab libraries.

Laboratory Specific Results:

Laboratory A: mitochondrial and microsatellite DNA

A total of 19 mtDNA control region haplotypes (excluding the sequential 5' 11 bp tandem repeat; see Faber and Stepien 1998) have been identified for walleye sampled from Lake Erie in our baseline data base of 324 walleye. Of the 19 haplotypes, only 4 (the most common haplotypes numbered 1-4 in Stepien and Faber 1998) were identified among the 200 walleye unknowns. Since the sites previously assayed (Maumee River and Van Buren Bay) each previously had rare haplotypes (totaling 12.5% for the Maumee River and 9.4% for Van Buren Bay from analyses of 1993-6 spawning runs), the genetic diversity in the unknown 2003 samples appears lower than previously observed. Haplotype #1 comprised 67 of the unknowns – equivalent to 33.5% of the individuals. Haplotype #2 characterized 55 individuals (27.5%), haplotype #3 was found in 30 individuals (15%), and haplotype #4 characterized 48 individuals (24%). Stepien and Faber (1998) hypothesized that these haplotypic spawning group frequency differences were due to differential glacial refugium recolonization patterns from the west (Mississippi refugium descendents) and the east (Atlantic refugium descendents) that may have been maintained by some spawning site philopatry.

In the present study, we then surveyed 3 microsatellite loci for the 200 walleye unknown individuals as well as 324 additional spawning site walleye (knowns). We have identified 11 alleles in the *Svi*4 locus (lengths 102-122 bp), 9 in *Svi*18 (118-134 bp), and 13 in *Svi*33 (80-104 bp) in our baseline data base from the walleye knowns. Of these, 9 of the *Svi*4 alleles were found in the walleye unknowns, 6 of the *Svi*18 alleles, and 11 of the *Svi*33 alleles. In all cases, including the mtDNA data, the 200 unknowns were

somewhat less diverse than is the entire data base of known walleye spawners from Lakes St. Clair and Erie.

Appendix A provides a comprehensive post-hoc analysis that uses variance in the “unknown” stocks, and in the baseline library, to identify the effects of different levels of aggregation on the overall correctness of the assignments.

Laboratory B: MH genes

The baseline set consisted of 51 Lake Erie fish typed for the MH class II alpha gene (total of 28 alleles), which was chosen because it had two alleles per animal and thus fit in the sequencing budget (MH class II beta is more polymorphic, but given there are up to 6 alleles per individual, it would have not been possible to isolate all the alleles for every individual within the budget provided). However in analyzing the 200 unknowns we found 51 alleles, many of them new. The results presented used the Goldstein distance method in the program Geneclass 2 because the small baseline data set made the use of more robust methods such as Bayesian techniques and the method used least gave an answer for all fish, with the exception of 3 individuals, for which no MH DNA could be isolated – most probably because they contained polymorphisms in the region where the primers should be binding. The program Structure (Pritchard et al. 2000) could not discriminate more than 1 population in the data from the 200 individuals without a previous assignment of the baseline individuals to a specific stock.

Given the small background dataset, which limited the analytical methods used to older techniques, MH typing performed reasonably well. A larger background dataset would substantially improve classification accuracy. Typing another locus such as MH class II beta would also provide an improvement in resolution.

Laboratory C: 2 separate microsatellite DNA multiplexes

To confirm that the unknown samples originated from Lake Erie, baseline data from other Great Lakes (Bay of Quinte (Lake Ontario); Ombabika Bay (Lake Nipigon); Nipigon Bay (Lake Superior) and associated populations (Georgia Lake and Lac des Mille Lacs, both inland stocking sources used for Nipigon Bay reintroductions) were also considered. Comparing assignment probabilities for the unknown samples against the probability distributions for reference genotypes from these outside populations resulted in the exclusion of the unknown genotypes as having originated from these potential external sources.

The microsatellite data highlighted the relatedness among stocks, and indicated past and ongoing straying and gene flow among breeding populations. Resolution of genetic structure among source populations in the baseline dataset was a tradeoff between analytical power and informativeness (stock discrimination). Self-assignment of reference genotypes from the baseline dataset was highly dependent on the degree of pooling among recognized spawning runs. When all potential Lake Erie source populations were considered as separate genetic entities, self-assignment of baseline data achieved only 38% success. Excluding Cattaraugus Creek (New York) and shoal-spawning fish from eastern Ontario waters, neither of which exhibited characteristics of cohesive breeding populations (data not shown), increased self-assignment resolution to 48% success. Pooling spawning runs into genetic groups that were statistically

different at a significance level of α = .05 (Grand River, Ontario; native New York populations (van Buren Bay and Smokes Creek); Hen Is.; other western basin populations) resulted in 68% of the baseline (reference) genotypes being correctly assigned back to their collection sites. Pooling the genetic data from multiple spawning populations resulted in significant loss of spatial resolution for identifying fish back to individual source populations, but substantially increased the percentage of fish that could be assigned back to their collection site for the source populations. Pooling the available baseline data into three genetic groups (Grand River, Ontario; native New York populations; and pooled western basin populations) resulted in 74% correct assignment. Assignment success of 95% was achieved for comparison of Grand River, Ontario walleye versus all other populations; however, this level of assignment power was only useful for directly assessing the contribution of Grand River fish to the set of unknowns, and was a prime example of the tradeoff between statistical power and spatial resolution (i.e. contribution of individual stocks for mixed-stock analysis). Comparable success was not achieved for any other single population, due to the genetic similarity at microsatellite loci for all other stocks considered.

Of the different hierarchical aggregations of baseline data from known spawning populations, pooling the reference genotypes into the three regional stocks (Grand River, ON; native New York; and western basin) gave the best compromise between maximizing assignment probabilities and retaining useful lake-wide geographic information. These groupings also reflect biological reality within Lake Erie, as all neutral genetic marker systems act as heritable “tags” that reflect the ancestry of individuals and populations.

Conclusions

This study has quantified the current state of stock discrimination for walleye in Lake Erie. However, the absolute results should not be taken too literally in that this study was designed more to identify for managers and for scientists what information gaps and expectations must be addressed before a mixed stock analysis can be applied to it’s truest potential. As stated in the introduction, genetic identity as applied to a natural population of fish is not as simple as reading a “bar-code”. Technologically, the science of genetic identity is highly advanced and is well suited to address management needs surrounding fish stock discrimination. However, certain biological realities exist that obscure individual signatures and / or our ability to assign fish to a “known” stock. These include the straying of individuals between stocks (mixing the underlying genetic signature) and / or stocking or fish transfers that lead to mixing of the underlying genetic signatures associated with a stock. Collectively, these factors lead to temporal instability across years, requiring frequent updates of libraries to ensure unknown fish are compared against the most probable current stock signature. In addition, a comprehensive library, representing all potential stocks of fish will increase the likelihood of defining discrete signatures that will more precisely assign the group of unknowns back to the correct stock. Finally, yet likely the greatest source of error in any mixed stock analysis, is our assumption that spawning aggregations of fish represent discrete stocks. If the correct samples are not collected initially to define a stock, then

any assignments made will be subject to these erroneous groupings. If the stock definition is too coarse (represents multiple genetic units) then the signature will become diffuse as it incorporates a unbalanced mix of different signatures. Conversely, if the stock definition is too fine (a single genetic unit divided among “stocks”) then the stock signatures will be little more than random groupings associated more with individual variation than with stock variation.

The different assignment outcomes from the different labs in part reflects differences in heritability and functionality among the different genetic marker systems that were used. Mitochondrial and microsatellite DNA markers are generally considered as neutral loci (i.e. not influenced by selection), but have different modes of inheritance and evolve at very different rates. By contrast, the importance MH genes in immune response and disease resistance may result in geographic structure resulting from local selective pressures. Each of these marker systems has its own strengths and weaknesses, but exhibited comparable levels of usefulness in this study. The lower predictive success from using MH genes largely reflects the much smaller baseline dataset that was available; it is highly probable that a larger baseline MH dataset would have resulted in much higher predictive success for the unknown samples. In order to make accurate assignments to individual stocks, it is probably best to couple different approaches and survey a larger number of loci.

Non-biological factors will also constrain management needs to apply mixed stock analyses. As stated above, stock assignments are only as good as the underlying reference libraries (known stock structure), so management must provide genetic laboratories with samples that represent, to the best of the management agency knowledge, complete representation of all discrete spawning groups (stocks). Costs associated with these studies appear high (Table 1), but given the recognised need for stock-specific management and the cost of other management initiatives (assessment of abundance, growth, diet, age, etc.) relative to the overall value of these fishery resources, enhanced management may offset this capital investment. The unit cost represents far more than just “running the sample” as the equipment and personnel capable of running these analyses are highly specialised, and therefore affiliated with academic institutions where overhead must be recouped (typically 30->50% on top of the project cost). Other costs are the obvious consumables, salaries, and maintenance and replacement of equipment. Given the limited number of laboratories equipped (trained personnel and hardware) to undertake these analyses, sample throughput (turn around time to deliver results) can be long. Because of the specialised nature of the work in a university setting, much of the labour (graduate students or technicians) is supported on soft money and may need to be sought out and hired specifically for these projects. Management must appreciate the costs and time lines for analyses inherent with the environment where this highly promising science is currently available.

We have summarised key recommendations for both managers and scientists in Table C1 to ensure that the expectations of each are aligned with technical realities of this rapidly emerging science.

Table C1.- Recommendations to maximise the success and discriminatory power of fisheries mixed stock analyses.

	Science Role	Management Role
Stock libraries	Stock libraries must be generated and maintained as a public (open-access) resource. Sharing of samples is not only cost-effective, it enhances agency appreciation for the pursuit of a common goal. Where possible, data collection tools (genetic data, morphometric data, ageing, etc.) should be standardized among participating labs to increase power of meta-analyses and decrease overall costs. Complementary structures (DNA, otoliths, digital images) should be housed together (GLFC coordinated central repository) or cross-referenced to facilitate comparison of stock identity and discriminatory power between techniques.	Fishery agencies must work closely with scientists who are developing / maintaining these libraries to ensure that samples represent stocks (discrete spawning units) or are identified within the library as fish of mixed origin.
	Libraries should be regularly updated to permit quantification of temporal stability in the underlying stock signatures.	Libraries should be regularly updated to permit quantification of temporal stability in the underlying stock signatures.
Communication	Scientists should take advantage of GLFC and agency meetings to communicate their findings to resource managers. Lack of exposure and misunderstanding of the hows and whys are the two primary communication challenges needed to build confidence and appreciation for the value of incorporating this science into management.	Managers should engage dialogue with scientists through active participation in meetings and workshops to ensure they remain apprised of the best available science to support management decisions.
Project Development	Scientists must work with management to develop an appropriate experimental design that will address both management needs and technical realities.	Managers must ensure their broader objectives are fully conveyed to the scientists (i.e. how will the information on stock structure be applied). This will ensure project outcomes are defensible.

Acknowledgements

We are indebted to agency staff with OMNR – Wheatley, ODW – Sandusky, and NYDEC for their assistance in collecting the walleye. Mike Lavender provided valuable assistance in extracting DNA. Mike Lavender, Kazuhiro Fujiki and Rex Strange assisted with DNA analysis and interpretation. This research was supported by a Restoration Act grant from the United States Fish and Wildlife Service. Bob Adair (USFWS) and Matt Engel (GLFC) assisted with the administration of the grant.

References

Banks, M.A. and W. Eichert. 2000. WHICHRUN (Version 3.2) a computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity. 91:87-89. 2004 (vers. 4.1) at: http://www.bml.ucdavis.edu/whichrun.htm

Bardakci, F. and D.O.F. Skibinski. 1994. Application of the RAPD technique in tilapia fish: species and subspecies identification. Heredity 73: 117-123.

Bartnik, S.E. 2005. Mixed stock analysis of a walleye cohort using otolith microchemisty. M.Sc. thesis. University of Windsor, Windsor, Ontario.

Berrill, M. 1997. *The Plundered Seas: can the world's fish be saved?.* Greystone Books, Vancouver.

Billington, N. and P.D.N. Hebert. 1988. Mitochondrial DNA variation in Great Lakes walleye (Stizostedion vitreum) populations. Can J. Fish. Aquat. Sci. 45: 643-654).

Borer, S., L.M. Miller, and A.R. Kapuscinski. 1999. Microsatellites in walleye *Stizostedion vitreum.* Molecular Ecology 8:336-338.

Burnham-Curtis, M.K. and G.R. Smith. 1994. Osteological evidence of genetic divergence of lake trout (*Salvelinus namaycush*) in Lake Superior. Copeia 4: 843-850.

Campana, S.E. 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. Mar. Ecol. Prog. Ser. 188: 263-297.

Campana, S.E. and J.M. Casselman. 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. 50: 1062-1083.

Campana, S.E. and S.R. Thorrold. 2001. Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? Can. J. Fish. Aquat. Sci. 58: 30-38.

Carvalho, G.R. and L. Hauser. 1994. Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology and Fisheries 4:326-350.

Carvalho, G.R. and T.J. Pitcher (eds). 1995. *Molecular Genetics in Fisheries*. Chapman and Hall, New York.

Cornuet, J.M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. Comparison of methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153: 1989–2000.

Dixon, B., L.A. Nagelkerke, F.A. Sibbing, E. Egberts, and R.J. Stet. 1996. Evolution of cyprinid major histocompatibility complex class II beta genes as reflected in the large barbels (*Barbus Intermedius* complex) of Lake Tana. Immunogenetics 44: 419-31.

Eldridge, W.H., M.D. Bacigalupi, I.R. Adelman, L.M. Miller, and A.R. Kapuscinski. 2002. Determination of relative survival of two stocked walleye populations and resident

natural-origin fish by microsatellite DNA parentage assignment. Can. J. Fish. Aquat. Sci. 59: 282-290.

Faber, J.E. and C.A. Stepien. 1998. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedion*. Molecular Phylogenetics and Evolution. 10(3):310-323.

Gatt, M.H., M.M. Ferguson, and A.P. Liskauskas. 2000. Comparison of control region sequencing and fragment RFLP analysis for resolving mitochondrial DNA variation and phylogenetic relationships among Great Lakes walleyes. Trans. Amer. Fish. Soc. 129: 1288-1299.

Gatt, M.H., T.L. McParland, L.C. Halyk, and M.M. Ferguson. 2003. Mitochondrial DNA variation and mixed-stock analysis of recreational and commercial walleye fisheries in Eastern Lake Erie. North Amer. J. Fish. Manage. 23: 431-440.

Goodyear, C.D., T.A. Edsall, D.M.O. Dempsey, G.D. Moss, and P.E. Polanski. 1982. Atlas of the spawning and nursery areas of Great Lakes fishes. Vol 1. Spawning and nursery areas of Great Lakes fishes: A summary by geographic area. US Fish and Wildlife Service, Washington, D.C. 166 pp.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Hatch, R.W., S.J. Nepszy, K.M. Muth, and C.T. Baker. 1987. Dynamics of recovery of the western Lake Erie walleye (*Stizostedion vitreum vitreum)* stock. Can. J. Fish. Aquat. Sci. 44: 15-22.

Hedges, K.J. 2002. Use of calcified structures for stock discrimination in Great Lakes walleye (*Stizostedion vitreum*). M.Sc. thesis. University of Windsor, Windsor, Ontario.

Hillis, D.M., B.K. Mable, and C. Moritz. 1996. Applications of Molecular Systematics: The State of the Field and a Look to the Future. Ch. 12, pp. 515-543 in: Molecular Systematics, 2nd Ed. Ed. D.M. Hillis, C. Moritz, and B.K. Mable. Sinauer Asoc., Inc., Sunderland, MA.

Hutchings, J.A. 2000. Collapse and recovery of marine fishes. Nature 406: 882-885.

Janeway, C., P. Traver, M. Walport, and M. Scholmchik. 2005. Immunobiology: The immune system in health and disease. 6th ed., Garland Science, New York and London.

Jarvis, R.S., H.F. Klodowski, and S.P. Sheldon. 1978. New mthod of quantifying scale shape and an application to stock identification in walleye (*Stizostedion vitreum vitreum*). Trans. Amer. Fish. Soc. 107: 528-524.

Knight, R.L. 1997. Successful interagency rehabilitation of Lake Erie walleye. Fisheries (Bethesda) 22: 16-17.

Krausse, M. 2002.New techniques for assessing growth and stock structure of percids in the Great Lakes. M.Sc. thesis. University of Windsor, Windsor, Ontario.

McParland, T.L., M.M. Ferguson, and A.P. Liskauskas. 1999. Genetic population structure and mixed-stock analysis of walleyes in the Lake Erie – Lake Huron corridor using allozyme and mitochondrial DNA markers. Trans. Amer. Fish. Soc. 128: 1055-1067.

Merker, R.J. and R.C. Woodruff. 1996. Molecular evidence for divergent breeding groups of walleye (*Stizostedion vitreum*) in tributaries of western Lake Erie. J. Great Lakes Res. 22:280-288.

Miller, K.M. and R.E. Withler. 1997. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. Hereditas 127: 83-95.
Miller, K.M., R.E. Withler, and T.D. Beacham. 1997. Molecular evolution at Mhc genes in two populations of chinook salmon *Oncorhynchus tshawytscha* Mol. Ecol. 6: 937-54.
Mion, J.B., R.A. Stein, and E.A. Marschall. 1998. River discharge drives survival of larval walleye. Ecol. Appl. 8: 88-103.
Olson, D.E. and W.J. Scidmore. 1962. Homing behaviour of spawning walleye. Trans. Amer. Fish. Soc. 91: 355-361.
Parham, P., K.L. Arnett, E.J. Adams, A.M. Little, K. Tees, L.D. Barber, S.G. Marsh, T. Ohta, T. Markow, and M.L. Petzl-Erler. 1997. Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. Tissue Antigens 50: 219-32.
Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GeneClass2: a software for genetic assignment and first generation migrants detection. J. Heredity (in press).
Post, J.R., M. Sullivan, S. Cox, N.P. Lester, C.J. Walters, E.A. Parkinson, A.J. Paul, L. Jackson, and B. J. Shuter. 2002. Canada's recreational fisheries: the invisible collapse? Fisheries (Bethesda): 27(1): 6-17.
Radtkey, R.R., B. Becker, R.D. Miller, R. Riblet, and T.J. Case. 1996. Variation and evolution of class I Mhc in sexual and parthenogenetic geckos. Proc. Royal Soc. London B Biol Sci 263: 1023-32.
Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248–249.
Regier, H.A., V.C. Applegate, and R.A. Ryder. 1969. The Ecology and management of walleye in western Lake Erie. Great Lakes Fish. Comm. Tech. Rep. No. 15., Ann Arbor, Michigan.
Riley, L.M. and R.F. Carline. 1982. Evaluation of scale shape for identification of walleye stocks from western Lake Erie. Trans. Amer. Fish. Soc. 111: 736-741.
Roseman, E.F., W.W. Taylor, D.B. Hayes, R.C. Haas, R.L. Knight, and K.O. Paxton. 1996. Walleye egg deposition and survival on reefs in western Lake Erie (USA). Ann. Zool. Fennici 33: 341-351.
Roughgarden, J. and F. Smith. 1996. Why fisheries collapse and whtat to do about it. Proc. National Acad. Sci. 93: 5078-5083.
Ryan, P.A., L.D. Witzel, J. Paine, M. Freeman, M. Hardy, S. Scholten, L. Sztramko, and R. MacGregor. 1999. Recent trends in fish populations in eastern Lake Erie in relation to changing lake trophic state and food web. p. 241-289 *in* M. Munawar, T. Edsall and I.F. Munawar (eds). *State of Lake Erie (SOLE) - Past, Present, and Future.* Backhuys Publishers, Leiden.
Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning (2nd edition). Cold Spring Harbor Laboratory Press, New York.
Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution 15: 199-203.
Schaefer, J.A. and C.C. Wilson. 2002. The fuzzy structure of populations. Canadian Journal of Zoology 80: 2235-2241.
Schneider S, Kueffer J, Roessli D and Excoffier L (2003) Arlequin vers. 2.2: A software

for population genetic data analysis. Univ. of Geneva, Switzerland: Genetics and Biometry Laboratory. http://lgb.unige.ch/arlequin/about.php

Smouse, P.E., and C. Chevillon. 1998. Analytical aspects of population-specific DNA fingerprinting for individuals. Journal of Heredity. 89:143-150.

Stepien, C.A. 1995. Population genetic divergence and geographic patterns from DNA sequences: Examples from marine and freshwater fishes. pp. 263-287. In Evolution and the aquatic ecosystem: Defining unique units in population conservation. Edited by J.L. Nielson. Am. Fish. Soc. 17. American Fisheries Society, Bethesda, MD.

Stepien, C.A., C.D. Taylor, and D.W. Einhouse. 2004. An analysis of genetic risk to a native spawning stock of walleye *Sander vitreus* (*Stizostedion vitreum)* due to stocking in Cattaraugus Creek. In *Proceedings of Percis III, the 3rd International Symposium on Percid Fishes*, T.P. Barry and J.A. Malison, eds. University of Wisconsin Sea Grant Institute, Madison, WI. 93-94.

Stepien, C.A. and J.E. Faber. 1998. Population genetic structure, phylogeography, and spawning philopatry in walleye (*Stizostedion vitreum*) from mtDNA control region sequences. Molecular Ecology. 7(12): 1757-1769.

Stepien, C.A. and T.D. Kocher. 1997. Molecules and Morphology in Studies of Fish Evolution. Ch. 1, pp. 1-11 In: Kocher, T.D. and C.A. Stepien (eds*.) Molecular Systematics of Fishes*. Academic Press, San Diego.

Stet, R.J.M., C.P. Kruiswijk, and B. Dixon. 2003. Major histocompatibility lineages and immune gene function in teleostean fishes: The road less taken. Critical Reviews in Immunology. 23(5&6): 441-471.

Strauss, R.E. and F.L. Bookstein. 1982. The truss: body form reconstruction in morphometrics. Syst. Zool. 31: 113-135.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution 15:199-203.

Todd, T.N. and R.C. Haas. 1993. Genetic and tagging evidence for movement of walleyes between Lake Erie and Lake St. Clair. Journal of Great Lakes Research 19: 445-452.

Walleye Task Group. 2004. Annual report of the Lake Erie Walleye Task Group to the Lake Erie Committee. Great Lakes Fishery Commission, Ann Arbor, MI.

Ward, R.D., N. Billington, and P.D.N. Hebert. 1989. Comparison of allozyme and mitochondrial DNA variation in populations of walleye, *Stizostedion vitreum*. Can. J. Fish. Aquat. Sci. 46: 2074-2084.

Waser, PM, and C. Strobeck. 1998. Genetic signatures of interpopulation dispersal. Trends in Ecology and Evolution. 13:43-44.

Wirth, T., R. Saint-Laurent, and L. Bernatchez. 1999. Isolation and characterization of microsatellite loci in the walleye (*Stizostedion vitreum*), and cross-species amplification within the family Percidae. Molecular Ecology 8: 1960–1962.

Wolfert, D.R. and H.D. vanMeter. 1978. Movement of walleyes tagged in eastern Lake Erie. New York Fish Game J. 25: 16-22.

Wright, J.M. and P. Bentzen. 1994. Microsatellites: Genetic markers for the future. Rev. Fish. Biol. Fish. 4:384-388.

Appendix A - Laboratory A Post-Hoc Test of Strength of DNA library

We tested the 200 unknown samples alone, assigning them to the 3 spawning population sites (after the identity of the unknowns were revealed to us). This served as a post-hoc control, for the use of the known baseline samples. We ran an addition post-hoc control with the 200 unknowns, assigning them to 2 populations – west (combining Maumee River and Hen Island) versus east (Van Buren Bay). Use of the known baseline data either then would improve or decrease the assignment accuracy.

Our baseline data set for the mtDNA control region, and 3 microsatellite loci sampled included 248 fish from 7 sites in Lake Erie. Sites included western Lake Erie spawning populations from the Maumee River, OH (N=32; 9 from 1994, 2 from 1995, 1 from 1998, and 20 from 1999); Sandusky River and Bay, OH (N=35; 14 from 1996, 1 from 1999, 20 from 2003); US west basin reefs, OH (N=15; all from 2003). Our west basin reefs did not include samples from Hen Island, Ont - as those samples were unavailable to us and confined to the Wilson and Dixon laboratories. We thus had to compare the Hen Island fish with the US west basin reef population, which may not be homologous. The central basin spawning sites were grouped together in all our analyses and included the Grand River, OH; Fairport, OH; and Chagrin River, OH; (total N=30; all from 2003). The eastern basin spawning sites in the baseline set included Van Buren Bay/Dunkirk, NY (N=50; 12 from 1994, 18 from 2001, 20 from 2003); Cattaraugus Creek, NY (N=55; 14 from 1998, 26 from 1999, 3 from 2000, 12 from 2003); and Grand River, Ont (N=31; 16 from 1995 and 15 from 2003). We conducted assignment tests of each unknown to the baseline using the 7 spawning sites. Once the spawning sites for the unknowns were revealed, our baseline thus included 4 spawning sites that were not included in the unknown locations, and lacked data for the Hen Island site. Our baseline thus included only 2 of the 3 spawning population sites for the unknowns, necessitating our including Hen Island unknowns with US west basin reefs. A further complication was that our baseline included several different spawning years, which may or may not show a close match with the 2003 spawners in the unknown sample, depending on spawning site fidelity and population genetic consistency among spawning year runs; which will be further tested in later analyses.

We then reduced the baseline known data set from 7 to 5 spawning population sites (to 187 individuals). We excluded 2 of our baseline sites; including the central basin spawning runs due to their relatively small contribution to the lake and small numbers of individuals, and the Grand River (Ont) spawning site due to its mixed origin from stocking from the Thames River, Lake St. Clair. This test was first run with only our known baseline spawning run samples (without the 200 unknowns). We next tested combinations of data (see Table A1), again using only our baseline data set (and without the 200 unknowns), assigned to 2 groups (west versus east) and then assigned to 3 groups (west basin rivers, west basin reefs, and east basin).

Next, we combined the 200 unknown samples (after known assignments were revealed) with our baseline data set and re-ran the analyses. We again ran these comparisons with the 7 spawning sites from our baseline data (including the central basin site and the Grand River, Ont), and Hen Island combined with our US west basin reefs. We then re-ran the tests using only 5 spawning sites (excluding our central basin spawners and the Grand River, Ont sample). Combinations of data were tested (see

Table A1) using our baseline data set and the 200 unknowns, assigned to 2 groups (west versus east) and then to 3 groups (west basin rivers, west basin reefs, and vs. eastern basin). The data sets also were run with only the 3 populations identified as sources (Maumee River, Hen Island/western reefs, and Van Buren Bay, adding 97 individuals to the 200 unknowns = 297). The identified 3 source populations were compared with combining Hen Island with our US west basin reef data and with using only Hen Island.

Assignment Tests:

The post-hoc self-assignment of each individual walleye to the 200 unknown data set (once the 3 spawning sites were identified) yielded 58.5% accuracy to the spawning sites and is represented as run A in Table A1, based on our 4 loci (1 mtDNA, 3 microsatellite). We also ran a post-hoc self-assignment of each individual walleye to 2 groups (west = Maumee River + Hen Island) versus east (Van Buren Bay). This self-assignment test yielded 70.5% accuracy to the basin and is shown as run B of Table A1. The condensing to west versus east analysis thus improved this prediction by 12%. We use these self-assignment tests to compare with our predictive data analysis runs.

The overall assignments with comparing the baseline spawning populations (numbering 7, 5, or 3 individual spawning runs) and grouping in 2 (west vs. east) or 3 groups (west basin rivers, west basin reefs, and east basin sites) thus may improve or decrease the overall resolution.

Using our known data base alone (excluding the 200 unknowns), the most accurate assignment was discerned with run F of Table A1, yielding 69.5% accuracy - which was an increase in 9% over the post-hoc self-assignment using only the unknowns to individual sites (run A) and a slight decrease of 1.0% from comparison with the unknowns alone to west vs. east spawning groups (run B). This assignment test involved 2 groups (west and east) and utilized 5 populations, including Maumee River, Sandusky River/Bay, US west basin reefs, Cattaraugus Creek, and Van Buren Bay. Thus both our unknown and our overall data sets showed relatively similar predictive assignment capacities, considering that the known data base had 187 individuals in it. Given similar sample sizes, the unknown versus known data bases yielded similar results.

In adding the 200 knowns and running the same known specimens (totaling 387 individuals), overall accuracy increased to 73.5% (run C of Table A1). This run had the greatest overall accuracy of all of our predictions (Tables A1 and A2). In this case, increasing the sample size increased the accuracy of assignment by 4% over the smaller sample size of knowns (without the unknowns), equivalent to 3% over the self-assignment prediction with 2 groups. Accuracy thus increases with increased sample sizes and increased numbers of sites within the west and east basins. The least accurate prediction of the 3 spawning sites occurred for Van Buren Bay (Table A2). The Van Buren Bay prediction was higher for the increased sample size (with the 200 unknowns) for run C than in run F. However, when one excludes spawning sites not used in the sampling design (as in runs D and E), one increases the overall prediction for the Van Buren Bay site alone, but not for the west basin sites (see Table A2).

The next- best overall predictions were found with run D of Table A1. This run again utilized 2 population groups (west and east) that included only Maumee River, Hen Island alone (without the US west basin reefs), and Van Buren Bay spawning individuals; and included all individuals sampled from those sites (both the 200 unknowns plus our baseline data of 97, totaling 397 individuals). This run thus excluded all non-target spawning sites (excluding 90 individuals used in run C). Table A2 shows that the greatest accuracy for run D was obtained with the Hen Island samples (81.8%), which was less than the accuracy with more sampling sites overall (compared with run C). This spawning run sample did not include the US west basin reefs in the prediction, which also were collected in 2003. In conclusion, adding the known baseline spawners from two of the sites improved the assignment prediction by 1.5% over the self-assignment level for 2 groups (run B), but did not improve the overall results compared to the inclusion of 5 spawning sites. This result was likely due to the decreased overall sample size. It did, however, improve the prediction for Van Buren Bay individuals alone – yielding the most accurate prediction for that site (Table A2). Since we did not have the Hen Island samples to include in our baseline study, this site had greatest accuracy without the west basin reef variability.

The third most accurate prediction occurred with analysis run E of Table A1, equivalent to 69.5% accuracy and 4% less than the self-assignment level using 2 groups (run B). This analysis again used 2 groups and the same criteria as in run D, but included the US west basin reefs from the known spawning site data. This inclusion did not improve the prediction, and probably increased the noise for the west basin reefs and Maumee River individuals. It specifically decreased the accuracy of correctly assigning the Hen Island population by 2.5% over run D, as shown in Table A2. Conversely, it matched the accuracy of correctly assigning the Van Buren Bay spawning individuals to the self-assignment test (run B) with 70.7%, improving the assignment by 3.5% over run C.

All other predictions were less informative. Additional noise was introduced by increasing the number of groups to 3 (including west basin reefs, along with west (Maumee + Sandusky) versus east (Cattaraugus Creek + Van Buren Bay) populations. Additional noise was also introduced by increasing the number of spawning site populations, i.e., including data from the central basin spawning sites and Grand River, Ontario. In conclusion, we do not at the present time have much confidence that our baseline data sets are large enough to resolve this problem with acceptable scientific accuracy and we need increased sample sizes and increased number of loci. It is clear from the results that increasing the number of loci and increasing the number of individuals are both highly desirable in order to achieve accurate predictions of stock assignments.

Table A1.- Results of correct assignments to spawning site or group, based on 4 loci (1 mtDNA, 3 microsatellite) by the Stepien and Strange Great Lakes Genetics Lab. ** = self-assignment test using 200 unknowns only *= best assignment results compared with self-assignment test. Note – Hen Island was lacking in baseline dataset and was merged with US reefs prior to analysis. With the exception of run N, no central basin sites, or the Grand River (Ont) were considered possible stocks. For group assignments, two groups represent west basin and east basin, 3 groups represent west basin reefs, west basin rivers, and the east basin.

Computer Analysis Run	Data source		Total N used to Assign	N sites in dataset	% Correct Assignment	N assignments		Stocks considered
	Used unknowns	Baseline library				Sites	Groups	
A**	Yes	No	200	3	58.5%	3		Maumee R, west basin reefs, Van Buren reef
B**	Yes	No	200	3	70.5%		2	West basin vs east basin
C*	Yes	Yes	387	5	73.5%		2	Maumee R, Sandusky R/Bay, US west basin reefs/Hen Is, Cattaraugus Cr, VanBuren reef
D*	Yes	Yes	272	3	72.0%		2	Maumee R, Hen Is, VanBuren reef
E*	Yes	Yes	297	4	69.5%		2	As in D with Hen Is & US west basin reefs grouped
F*	No	Yes	187	5	67.5%		2	As in C using known baseline library only
G	Yes	Yes	297	4	54.5%	3		As in E
H	Yes	Yes	272	3	53.0%	3		Maumee R, Hen Is, Van Buren reef
I	Yes	Yes	387	6	48.5%		3	As in C
J	Yes	Yes	248	2	48.5%	2		Maumee R & Van Buren reef only
K	Yes	Yes	387	6	47.5%	5		As in C
L	No	Yes	187	5	44.0%		3	As in F
M	No	Yes	187	5	34.0%	5		As in F
N	No	Yes	448	7	14.5%	7		Used central basin and Grand River (Ont) spawning sites, in addition to C

Table A2.- Breakdown results of correct assignments to the three spawning sites for the post-hoc analysis conducted by the Stepien and Strange Great Lakes Genetics Lab.

Computer Analysis Run (see Table 1)	*Maumee River*	*Hen Island*	*Van Buren Bay*	*Overall*
B	67.1%	74.2%	70.7%	70.5%
C	82.8%	87.9%	46.6%	73.5%
D	67.1%	81.8%	67.2%	72.0%
E	68.4%	68.2%	70.7%	69.5%
F	80.3%	80.3%	34.5%	67.5%