DNA Barcoding and Related Molecular Markers for Fish Species Authentication, Phylogenetic Assessment and Population Studies

by

Li Lian Wong

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2011

Copyright 2011 by Li Lian Wong

Approved by

Zhanjiang Liu, Chair, Professor of Fisheries and Allied Aquacultures
Michael C Wooten, Professor of Biological Sciences
Nannan Liu, Faculty Endowed Professor of Entomology and Plant Pathology
Eric J Peatman, Assistant Professor of Fisheries and Allied Aquacultures
Huseyin Kucuktas, Research Fellow IV Fisheries and Allied Aquacultures

Abstract

This study highlighted the usage of molecular genetic markers in fish species and population level studies to address food safety and ecological issues. Species level study was constructed based on the growing concern in the food safety arena involving seafood fraud. The first species level study focused on the validation of DNA barcoding using cytochrome oxidase I (COI) gene for catfish species discrimination. 651 bp barcodes from 9 catfish species (and an Ictalurid hybrid) represented by families of Ictaluriidae, Clariidae, and Pangasiidae were generated. Most of the catfish species consensus barcodes constructed from sequence alignments were in agreement with the recorded sequences in two major databases (GenBank and Barcode of Life Data Systems). Validation tests carried out in blinded studies also revealed the reliability of DNA barcoding in species identification. The second study involved the development of microchip-based Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) to be compared with DNA barcoding for efficiency in catfish species identification. Results found that RFLP profiles derived from a combination of all three enzymes were able to produce consistent restriction fragment patterns for all catfish species. Compared to conventional gel-based electrophoresis, this sensitive and robust endpoint detection method can increase resolution of RFLP patterns. In comparison, DNA barcoding has a lower operation cost and longer data collection time, whereas lab-on-chip PCR-RFLP is more suitable for a small scale and rapid identification of catfish species. Finally, the last study focused on the population structure of redeye bass (*Micropterus coosae*) populations in three main water drainages in Alabama using 10 polymorphic microsatellite loci. Substantial genetic differentiation was

observed among three major river basins with the highest divergence observed between Coosa Basin and Cahaba Basin (Fst=0.256), indicating restricted gene flow at the micro-geographical scale. Both STRUCTURE analysis and a UPGMA phylogenetic tree revealed four genetic clusters with significant admixture events between water drainages. Overall, the pattern of genetic structuring, strongly associated with river basins rather than geographic distance, signified the occurrence of human-mediated translocations.

Acknowledgments

I would like to express my deepest gratitude and sincere appreciation to my major professor, Dr. Zhanjiang Liu, for his guidance throughout my graduate study. I would also like to thank my advisory committee: Dr. Eric Peatman, Dr. Huseyin Kucuktas, Dr. Michael Wooten and Dr. Nannan Liu for their advice and guidance in completing my dissertation. My thanks also go to all my colleagues in the laboratory for their help, support and collaboration. My last gratitude goes to my parents and family members for their moral support and encouragement.

Table of Contents

Abstract ii

Acknowledgments iv

List of Tables vii

List of Figures ix

Introduction 1

Chapter 1 DNA Barcoding of Catfish: Species Authentication and Phylogenetic Assessment ... 22

Abstract 23

Background 24

Materials and Methods 26

Results 29

Discussion 33

References 38

Chapter 2 Catfish Species Identification Using Lab-On-Chip Polymerase Chain Reaction Restriction Fragment Length Polymorphism (PCR- RFLP) 49

Abstract 50

Background 51

Materials and Methods 53

Results and Discussion 56

Conclusion 62

References 63

Chapter 3 Population Structure of Redeye Bass (*Micropterus coosae*) in Alabama 72

Abstract 73
Background 74
Materials and Methods 76
Results 80
Discussion 83
References 87
General Discussions and Conclusions 97
References 100

List of Tables

Table 1 Catfish species used in this study 41

Table 2 Summary of identification based on each species consensus barcoded sequence using BOLD Identification System (BOLD-IDS) and BLASTN search from GenBank.....42

Table 3 Description of analyzed local market samples 43

Table 4 Estimates of Pairwise Genetic Distances between Catfish Species under Kimura Parameter Model 44

Table 5 Catfish species used in the development of PCR-RLP profiles.....65

Table 6 Predicted and observed PCR-RFLP fragment sizes generated with three different restriction endonucleases *Dde*I, *Hae*III, and *Nla*III of catfish species 66

Table 7 Summary of identification matches based on each catfish species consensus sequence of cytochrome b region BLASTED against entries in GenBank database 67

Table 8 Sample sizes and summary of genetic diversity parameters of the redeye bass at 10 microsatellite loci sampled in Alabama**.**.....90

Table 9 Pairwise Fst estimates (below diagonal) and their P-values (above diagonal) among all locations 91

Table 10 Analysis of molecular variance (AMOVA) based on Rst distance method comparing the genetic variation in 10 microsatellite loci for *Micropterus coosae* from three main water drainages 92

Table S1 Primers used for PCR amplification and sequencing……45

Table S2 Reference sequence numbers (BOLD) and accession numbers (GenBank) of voucher species used to build multiple sequence alignment of *Ictalurus punctatus* using CLUSTALW program 46

Table S3 Microsatellite primer sequences, PCR conditions, amplicon sizes and sources used in this study for *Micropterus coosae* populations in Alabama. The locus name listed included both forward and reverse primer sequences.....96

List of Figures

Figure 1 Phylogenetic consensus tree of nine catfish species (and an Ictalurid hybrid) constructed using Maximum Parsimony (MP) Method 47

Figure 2 Multiple sequence alignment of consensus sequences for *Ictalurus punctatus* against voucher and reference sequences from GenBank and BOLD databases.....48

Figure 3 Computer generated Restriction Fragment Length Polymorphism (RFLP) gel image on 2100 Bioanalyzer for cytochrome b fragments of catfish species 68

Figure 4 Electropherograms show the presence of species defining peaks (band size) for *Ictalurus punctatus* which corresponds to the RFLP profiles generated by three restriction endonucleases 69

Figure 5 Species identification flowchart developed for 7 catfish species (and an Ictalurid hybrid) based on the PCR-RFLP profiles generated by 2100 Bioanalyzer.....70

Figure 6 Phylogenetic consensus tree of seven catfish species (and an Ictalurid hybrid) constructed using Maximum Parsimony (MP) Method.....71

Figure 7 Schematic representation of sampling sites and letter codes of redeye bass populations collected from three river basins.....93

Figure 8 Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic tree reconstructed using MEGA 5.0 based pairwise Fst 94

Figure 9 Summary plots of the estimated membership coefficient for each redeye bass individual from all populations based on 10 microsatellite loci 95

INTRODUCTION

Difficulties arose when morphology-based identification is used to diagnose samples lacking morphological characteristics such as fish fillets and processed fish products (Sotelo et al. 1992; Unlusayin et al. 2001; Smith et al. 2008). Although there are numerous protein-based analyses that have been developed specifically for fish species identification (Tepedino et al. 2001; Ochiai et al. 2001), these methods are only reliable for identification of raw fish. They are not applicable for processed fish products because thermally labile proteins of fish will become irreversibly denatured by heat (Dooley et al. 2005). Furthermore, differentiation of closely related fish species is difficult as they share a common set of protein profiles (Bartlett and Davidson 1991; Smith et al.1996). Obviously, when differentiation of species is difficult, determination of fish population structure is even more difficult because the differences among populations of the same species are substantially minor when compared with differences among species. Therefore, alternative DNA-based techniques with high thermal stability as well as being cell and age independent should be employed to enable fish species identification, phylogenetic assessment, and population genetic studies (Davidson 1998; Bossier 1999; Lockley and Bardsley 2000).

Manifold of DNA-based methods which are categorized into mitochondrial and nuclear DNA markers have been widely applied in seafood and livestock authentication as well as in population genetics and molecular phylogenetics studies (Lockley and Bardsley 2000; Arif and Khan 2009). Those specific molecular markers include Restriction Fragment Length Polymorphism (RFLP) (Cespedes et al. 1998; Hold et al. 2001; Sanjuan and Comesana 2002), Single Strand Conformation Polymorphism (SSCP) (Cespedes et al. 1999), Random Amplified Polymorphic DNA (RAPD) (Asensio et al. 2002), Amplified Fragment Length Polymorphisms

(AFLPs) (Watanabe et al. 2004; Maldini et al. 2006; Gonzales Fortes et al. 2008) and microsatellites (Banhos et al. 2008; Kawaka et al. 2007).

Fish Species Identification

The growing demand of seafood products has provoked economic deception involving cases of fish species misbranding. Incidence of seafood frauds involving substitution or mislabeling of high-end market fish with others of lower-priced is common (Rasmussen et al. 2009; Hsieh et al. 1995; Miller and Mariani 2010). In addition, this commercial mislabeling has also been reported to raise food safety concern in the case of mislabeled toxic puffer fish as "headless monkfish" (Cohen et al. 2009). Moreover, fisheries conservation sustainability has also been neglected to the point of commercializing overexploited species (Jacquet and Pauly 2008). In view of this, seafood frauds are most prevalent in processed fish products or fish fillets when they could not be readily tractable due to the removal of morphological features (Marko et al. 2004; Wong and Hanner 2008). In addition, substitution of closely related species from other countries or continent is also concurrent with the discrepancy of the commercialized fish nomenclature. Here, some species are labeled with only a single name such as in the tuna species (Lowenstein et al. 2009) while the other species are given multiple names (Barbuto et al. 2010).

All these fraudulent cases have also occurred in catfish species. Catfish, specifically the Ictalurid species dominates the domestic aquaculture industry in the United States (USDA 2002). At the same time, imports of Pangasiid, Clariid, and Ictalurid catfish to the United States from East Asia (largely Vietnam and China) have also increased rapidly over the last decade and now account for up to half of catfish sales in the USA (Hanson 2009). However, anecdotal and documented cases of catfish species mislabeling (either as another catfish species or as a higher

value species) have also jeopardized the socio-economy of US catfish industry. In fact, pollution in farming environment of these imported catfish has also become a highly concerned public health issue (Brambilla et al. 2009). Thus, accurate catfish species labeling is essential in aiding consumers towards making informed purchasing decisions (Roheim and Sutinen 2006; Heyden et al. 2010). To protect consumers from market deception, new regulations are currently under development by the federal government to strengthen inspection of domestic and imported catfish, including verification of correct species labeling. Hence, the development of innovative, rapid and low cost analytical platforms for species identification is of high importance to guarantee the authenticity and origin traceability of catfish along the seafood distribution chain from pond to plate.

The first two chapters in my dissertation focused on comparing the efficiency of two different molecular genetics platforms in catfish species identification. The first species identification method involved the test and validation of DNA barcode techniques for domestic and imported catfish, including the creation of a DNA barcode database containing eight worldwide-commercialized catfish as well as two wild populations of catfish species. Species involved in this study (Table 1) included channel catfish (*Ictalurus punctatus*), blue catfish (*Ictalurus furcatus*) and hybrids of channel (♀) x blue (♂) catfish; Southeast Asian walking catfish (*Clarias batrachus*), broadhead catfish (*Clarias macrocephalus*), basa (*Pangasius bocourti*), swai or sutchi (*Pangasianodon hypopthalmus*); and African sharp-toothed catfish (*Clarias gariepinus*). To complement this study for the purpose of detecting market substitution with vulnerable species, two wild catfish species from China were also included: helmet catfish (*Cranoglanis bouderius*) and long-barbel catfish (*Hemibagrus macropterus*). On the other hand, the second identification tool involved the collaboration with Agilent Technologies in

developing PCR-RFLP profiles on the 2100 Bioanalyzer to authenticate the similar catfish species as in the first study but excluding specimens of the Chinese catfish species.

Catfish species under study

Catfish (Order Siluriformes) are diverse groups of fish with constitution of more than 3,000 species, 478 genera and 36 families (Ferraris and de Pinna 1999, Eschmeyer et al. 2004). Over the last 50 years, catfish farming in the USA has developed rapidly and became the largest producers (47%) in the commercial finfish industry (FAO 2002), generating an approximately 500 million pounds of catfish per year (USDA 2002). However, the industry has suffered an intense competition with imported catfish from other countries, which may hamper the current growth or even cause the decline of the market level. The non-domestic catfish are claimed to be exotic catfish from Vietnam as well as channel catfish grown in China (FAO 2006).

Air-breathing catfish from the family Clariidae is composed of 15 genera and 93 species, distributed in freshwaters of Africa, Asia Minor and Southeast Asia (Teugels et al. 2001; Teugels and Adriaens 2003; Mohindra 2007). They are normally characterized by their unique suprabranchial organs (Teugels and Adriaens 2003), which enable them to respirate atmospheric oxygen (Greenwood 1961; Graham 1997). Their air-breathing capacity in various hypoxic environments (Matthes 1964; Burgess 1989) places these fishes as promising species with great economic potential in aquaculture (Mohindra 2007). Members of family Clariidae that are commercially important are *Clarias batrachus* (L.) and *Clarias macrocephalus* Gu¨nther (Na-Nakorn 1999) which are native to the Indian subcontinent and Southeast Asia whereas African sharp-toothed catfish, *Clarias gariepinus* (Burchell) (Proteau 1996) are introduced for cultivation in Asia.

Catfish family of Pangasiidae is consist of over 20 species and is native to the freshwater in South East Asia (Pouyard 2000). Pangasiid catfish have been wild captured as food fish and is currently gaining popularity in the cultivation industry (Hogan and Bernie 2002). The export value of *Pangasianodon hypophthalmus* or Swai catfish has reached up to USD 1.6 billion in 2008 which has flourished the economy of the producing country, Vietnam (FAO 2006).

Of the two wild catfish species in the present study, one of them, helmet catfish (C*ranoglanis bouderiusis*) has been described as a closely related sister taxa to the North American catfish family of Ictaluriidae (Hardman 2005). This species is endemic to the Pearl River, Red River and drainages in Hainan Island. It was a food fish in the Pearl River Basin before being enlisted as a vulnerably endangered species in the Chinese Red Data Book of Endangered Animals (Yue and Chen 1998). On the other hand, *Hemibagrus macropterus* (Bleeker 1870) is a catfish member of the family Bagridae native to mainland China, and is mainly distributed in Yangtze River, Changjiang (Chu et al. 1999).

Mitochondrial DNA and DNA barcoding

Mitochondrial genes are promising markers for fish species identification (Kochzius 2009; Teletchea 2009) when compared to nuclear genes because of its several special features. Mitochondrial DNA (mtDNA) emerges in high copy number in each cell as a small closed circular DNA with the size range of 15-20kb, which could ease the recovery of DNA with various extraction methods (Hubert 2008). Its maternal inheritance pattern without recombination (Sangthong 2003), rapid mutation rate and small effective population make it a decent tool for studying phylogeny and genealogy of taxa through matrilineage (Moore 1995). Additionally, the evolution rate of mitochondrial genes is fairly rapid and therefore exhibits a

great potential as barcoding genes. Mitochondrial genes such as cytochrome b (cyt b) and 16S rRNA are common fish species identification markers that have been widely utilized in seafood control (Quinteiro et al. 1998; Sotelo et al. 2001; Santaclara 2006), fisheries control (Greig et al. 2005; Hoelzel 2001), and species delineation (Kochzius et al. 2003; Lemer et al. 2007). Due to these beneficial attributes, mitochondrial genes have been used extensively in molecular tagging or DNA barcoding. However, mitochondrial markers also have their limitations: 1) the mitochondrial (mt) genome is small, most often just about 15-20 kb in size in fish. As compared with nuclear genomes, it is a small proportion of the genetic material; 2) mt markers are maternally inherited, which on the one hand allows tracing of maternal contributions, but on the other hand, the non-Mendelian inheritance makes it unsuited to many genetic studies; 3) Due to the high mutation rate and small size of mt genomes, back mutation can readily happen that does not reflect the phylogenetic relationship or evolutionary history.

The decreasing numbers of taxonomist and the limitations of morphological-based identification increase the need for a more rapid approach in taxon delimitation. Microgenomic identification systems, which adopt the usage of small segments of the genome for the discrimination of diverse forms or taxa, have birthed the DNA barcoding concept. While traditional taxanomic analysis involves the collection of morphological and ecological data that are different for each taxanomic assemblages and species, DNA barcoding represents a more standardized method that can be applied across a broader swath of living taxa.

DNA barcoding is based on the principle of sequencing a short segment of DNA from a uniform region of the mt genome of target specimen and comparing these unknown barcodes to an existing barcode database to identify species (Hebert et al. 2003). Most often, cytochrome oxidase I or cytochrome b sequences are used. DNA barcoding is also known to refine species

discovery by identifying query specimens using probabilistic algorithms when a set of barcodes of known species is established (Abdo and Golding 2007; Nielsen and Matz 2006). Recent results have demonstrated significant benefits from the usage of DNA barcoding as a standardized molecular tool for species identification (Hebert et al. 2003; Hebert and Gregory 2005). DNA barcoding has been widely employed in various non-physically tractable organisms (Nanney 1982; Pace 1997) and has since exploited the diversity of barcode sequences (Kurtzman 1994; Wilson 1995) that are found in every cell, for species identification.

As mentioned above, DNA barcoding identify a species from short standardized fragments that can be applied to different kinds of specimens from processed foodstuffs to fossil samples, as well as various stages of life for a particular species. Besides, DNA barcoding also provides an unambiguous digital identification mode in the form of four discrete nucleotides, in contrast to the less conclusive analog (e.g shapes and colors) morphological description of a species. Although a barcode gene might not be effective in phylogenetic tree reconstruction, it should still be helpful for revealing the speciation process by comparing closely related organisms. Most importantly, a DNA-based identification system is more rapid and informative than the traditional morphological-based methods and can be a crucial tool especially in highly critical field of conservation biology (DeSalle and Amato 2004).

An ideal barcoding gene should be present in most forms of life and its minimal sequence length should be sufficient enough for species level discrimination. This short fragment will also ease DNA recovery from degraded specimens as well as aiding the development of chip-based DNA arrays. In order to develop a uniform animal species identification system, the adoption of global standard in DNA barcoding has been advocated through the usage of 650 bp fragment of the 5’ end of the mitochondrial gene cytochrome oxidase I (COI) region (Hebert et al. 2003).

COI region has not only been reported to be able to distinguish closely related species (Hebert et al. 2003) in previous studies of both marine and freshwater fishes (Steinke et al. 2005; Ward et al. 2005), but has been widely employed in various biological fields ranging from forensic sciences (Dawnay et al. 2007), molecular systematics (Hardman 2005) to seafood products identification (Botti and Giuffra 2010; Lowenstein et al. 2010; Steinke 2009). Certainly, there are several advantages that lead to the usage of COI as the barcoding gene. First of all, COI region is one of the protein coding genes in mitochondrial genome besides cytochrome b which is abundant in all eukaryotes. Moreover, the availability of robust broad-range universal primers also allows the recovery of this gene region from diverse unknown specimens of animal phyla (Folmer et al. 1994; Zhang and Hewitt 1997). COI also outperforms the other mitochondrial gene with a greater range of phylogenetic signal. Its higher rate of base substitution in its third-position nucleotides in codon makes it possess an evolutionary rate of three times greater than that of 12S or 16S rDNA (Knowlton and Weigt 1998). Because of its rapid evolution rate, this gene also enables the differentiation of closely related species as well as phylogeographic groups within a single species (Cox and Hebert 2001; Wares and Cunningham 2001). In fact, COI can provide a higher level of resolution than popularly used mitochondrial gene such as cytochrome b (Simmons and Weller 2001) due to the rapid changes in its amino-acid sequence (Lynch and Jarrell 1993). Besides, the alignment of COI sequences is uncomplicated because the existence of INDELS in most mitochondrial genes is rare (Mardulyn and Whitfield 1999). Its small fragment size is also readily obtainable by a single sequencing direction in conventional cycle sequencing platforms. In addition, shorter fragments of COI region have also been reported to be effective in identification of suboptimal specimens especially of those with degraded DNA or when obtaining full-length barcode is impossible (Hajibabaei et al. 2006). Moreover, it is also

documented that single nucleotide sequence as short as ~50 bp of 5' region of COI barcode contains sufficient information to generate unique character-based DNA diagnostics for mammalian species (Hajibabaei 2007).

The robustness of COI in fish species discrimination has led to the formation of international fish barcoding program (FISHBOL; www.fishbol.org) (Kochzius 2009; Ward et al. 2009). Out of almost 30,000 fish species estimated in the world, barcodes for more than 23,000 fish species are currently recorded in the Barcode of Life Data Systems (BOLD) database. BOLD is a community-based effort that is focused on developing extensive DNA barcode libraries, and most notably has led to the adoption of DNA barcoding technology as the gold standard for species identification. This database also provides detailed information of COI-sequenced species including the origin and current location of voucher specimens (Ratnasingham and Hebert 2007). In addition to barcoding, other morphological, behavioral, ecological, natural history and geographic variation data should be used in complimentary with the sequence information. Thus, the integration between DNA barcoding and the conventional phenotypic approach allows a higher resolution in species discrimination across diverse animal phyla (Ward et al. 2005; Ferri et al. 2009). In cases of processed fish, barcoding may be a reliable approach as morphological information is no longer available. I will discuss further the application of DNA barcoding using COI gene on catfish species identification in Chapter 1 of my dissertation.

Lab-on-chip PCR-RFLP

Although DNA-based methods such as real-time PCR and direct sequencing provide definitive species identification, their higher cost and time-consuming process, respectively make them less suitable for a robust and routine fish-screening test (Meyer et al.1995). DNA barcoding, an

approach that is effective, requires sequencing of the specimens. Therefore, a relatively simple and inexpensive species discrimination technique such as PCR-RFLP should be developed to compare and determine the most appropriate method for species identification. In fact, PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) mapping have been successfully applied in various fish species and fish products identification (Arahishi 2005; Akasaki et al. 2006; Lin and Hwang 2007; Cespedes et al. 1998; Hsieh et al. 2007). Briefly, this technique incorporates the analysis of specific DNA fragment profiles generated by digestion of PCR amplicons with selected restriction endonucleases to produce restriction fragment patterns through electrophoretic sorting. However, PCR-RFLP still relies on conventional gel electrophoresis for endpoint detection, which is technically hazardous, time demanding and complicated by the production of variable results (Dooley et al. 2005). The complex DNA fingerprints and sizing variation has led to the development of micro-chip based capillary electrophoresis technology such as automated lab-on-chip electrophoretic system replacing the traditional gel electrophoresis step (Fajardo et al. 2009; Panaro et al. 2000). The Agilent 2100 Bioanalyzer is the first commercialized chip-based electrophoresis apparatus that utilize capillary electrophoresis (CE) technology. This automated system involves DNA fragments separation by CE in miniaturized channels of the microfluidic chip followed by detection steps using laser-induced fluorescence (Vasilyeva et al. 2004). This user-friendly technology offers advantages of speed and safety, permitting the generation of higher resolution PCR-RFLP profiles in comparison to conventional gel-based approaches (Uthayakumaran et al. 2005). Though the overall cost of this advance device is fairly high, it is still ideally suited for rapid and routine surveys of smaller sample sizes compared to the costly real-time PCR method (Dooley and

Garrett 2001). In Chapter 2 of my dissertation, I will describe my work on the development and application of lab-on-chip PCR-RFLP for the identification of catfish species.

Fish Population Structure Studies

Population structure and differentiation of freshwater fish in river systems are normally concordant to the isolation by distance model (Meffe and Vrijenhoek 1988; Thuesen et al. 2008). Due to the shared dependence of freshwater for fish and humans, the decline of fish population size as well as the increased number of fish species enlisted in the "IUCN Red List of The Threatened Species" is not unexpected (Duncan and Lockwood 2001). Historically and genetically related fish populations within water drainages maybe structured by natural or artificial barriers, dispersal propensity, habitat alterations and fishing pressure resulting in changes of gene flow pattern which impedes fish dispersal (Yamamoto et al. 2004; Wofford et al. 2005). As a result of low connectivity between fish populations, a high level of differentiation between populations can be generated (Hughes 2007).

In view of this detrimental environmental issue, redeye bass (*Micropterus coosae*), an endemic centrachid of Southeastern United States is not an exception either, and is currently becoming imperiled due to drainage alteration, aquatic pollution and invasion of exotic species. The rising popularity of redeye bass in recreational fisheries has also increased its human-mediated introduction into previously uninhabited watersheds (Rohde et al. 1994). This will also create genetic pollution via hybridization with the native redeye bass species, which adversely affect the genetic function of the new ecosystem. On the other hand, genetic assimilation also occurs via hybridization with invasive congeners (other black bass species), which will extirpate native species to extinction (Rubidge and Taylor 2004). Genetic assimilation is an eventual

extinction of a natural species driven by excessive gene flow from another related species, causing the natural species to become a new species that are characterized by distinct genomes. In addition, fragmentation of watershed will also prevent migration of redeye bass populations, thus reducing gene flow and increasing divergence among populations. Drainage impoundment or dam constructions act as artificial barrier for redeye bass migration, restricting gene flow among populations and eventually cause populations differentiation (Ward and Stanford 1983).

Genetic diversity is vital for both short-term fitness of individuals and long-term survival of populations (Ferguson et al. 1995). Therefore population structure assessment is fundamental for the maintenance of intra-specific genetic diversity and future fisheries management and conservation of redeye bass (Nelson and Soule 1987). Currently, stocking or re-introduction is a popular and extensively used strategy to replenish extirpated fish populations. However, concerns about the possible genetic impacts on the remnant fish populations from this program have been recently raised (Butler 2009). In the absence of genetic structure information, the anthropogenic stocking of non-native stocks from locations with different levels of genetic diversity could disrupt the natural patterns of genetic variation. Besides, in some cases this will lead to the extent of permanent loss of genetic diversity, specifically when the source and recipient populations are genetically distinct (Englbrecht et al. 2002; Nelson and Soule 1987; Echelle 1991). Most importantly, this controversial strategy could result in detrimental effects such as depression, hybridization and parasite introduction, which could potentially reduce the fitness of populations (Moritz 1999; McClelland and Naish 2007). Therefore, a comprehensive study on the population structure and differentiation of redeye bass populations is crucial for the development of a conservation and restoration plan. For an effective plan, documentation of consistently scorable alleles of polymorphic markers as well as pattern and distribution of

genetic diversity of targeted species are necessary (Ferguson et al. 1995). In Chapter 3 of my dissertation, I will provide insight into the population genetic pattern of redeye bass in the main water drainages in Alabama using 10 polymorphic microsatellite loci. Parameters of genetic differentiation computed by this dataset will be used to determine the levels of genetic divergence.

Redeye bass

The redeye bass (*Micropterus coosae*) is a species of black bass endemic to the Southeastern (SE) United States and are typically found in cool streams and rivers in the foothills of mountains (Koppelman and Garrett 2002). Redeye bass are restricted to a few drainages within the SE region, ranging from Black Warrior, Cahaba, Coosa and Tallapoosa of the Mobile River basin and tributaries of the Alabama River in Alabama (Hurst et al. 1975) to the headwaters of Savannah River upstream of Hartwell Reservoir in South and North Carolina (MacCrimmon and Robbins 1975). Redeye bass are also indigenous to the Alabama River system and Chattahoochee River (Apalachicola River drainage) above Columbus (Pipas and Bulow 1998) in Georgia. Other *Micropterus sp.* that are distributed in the SE United States are shoal bass (*M. cataractae),* smallmouth bass (*M. dolomieu*), spotted bass (*M. punctulatus),* largemouth bass *(M. salmoides)* and Guadalupe bass (*M. treculii).*

Redeye bass have high aesthetic value (Hurst et al. 1975), distinguishable from other black basses by its color characteristics. They have brick red eyes and fins as well as deep bronze, greenish or purplish back which changes according to the habitat background (Parson 1954). Redeye bass also resemble smallmouth bass physically but are taxonomically associated with spotted bass (Hubbs and Bailey 1940). Redeye bass are readily distinguishable from shoal

bass and smallmouth bass by the appearance of the white edge on the lower and upper part of caudal fin. The young of redeye bass does not have sub-terminal black band across the caudal lobes like those found in other basses (Parson 1954).

Redeye bass have a slower growth rate compared to other black basses (Hurst et al. 1975). These fish mostly feed on terrestrial and aquatic insects (Tatum 1965; Cathey 1973; Gwinner 1973). Crayfish and small fishes are also source of food for redeye bass although not as popular (Hurst et al. 1975). Spawning season is during May and June in coarse gravel beds with temperature ranging from 17-21°C (Parson 1954). Redeye bass become sexually mature at 3-4 years of age, depending on the stream conditions (Hurst et al. 1975). They coexist with largemouth and spotted bass and successfully sustain themselves with other black basses only if the growth of other bass species is slow. Redeye bass are aggressive fighters and jump frequently when caught by anglers (Hurst et al. 1975).

Anthropogenic Introduction, Genetic assimilation and Habitat alteration

Redeye bass possess a complex demographic history to sustain their genetic diversity in the Appalachian headwaters of SE United States (Boschung and Mayden 2004). This fish become increasingly popular for recreational fishing, which exposes it to human-mediated translocations to watersheds outside its native range. Rohde et al. (1994) reported several anthropogenic introductions of icthyofaunal species within South Carolina drainage system. This supports Koppelman and Garrett's (2002) hypothesis that redeye bass populations in the Saluda River consist of both native and anthropogenically translocated individuals. Another anthropogenic introduction of redeye bass occurred in the early 1960's in north-central Tennessee streams (Pipas and Bulow 1998). Other species of black bass were also illegally translocated for sport

fishery enhancement. For example, non-native Alabama spotted bass (*Micropterus henshalli*) were introduced into a native redeye bass drainage, Savannah River's Keowee Reservoir in the early 1980's (Barwick and Moore 1983).

Subsequent to this introduction of other black bass species, annual reservoir fishery surveys discovered the decline of redeye bass abundance and an increment of hybrid fishes with morphologic characteristics of both Alabama spotted bass and redeye bass (Barwick and Moore 1983) as well as smallmouth bass and redeye bass (Pipas and Bulow 1998). Barwick et al. (2006) furthered confirmed the presence of redeye bass and Alabama spotted bass hybrids within Keowee Reservoir through genetic studies. Moreover, report shows that the genus *Micropterus* (black bass) has high propensity for hybridization and the hybrids are often fertile (Philipp 1991), with the ability to rapidly propagate and colonize the new aquatic habitats within short time periods (Avise et al. 1997). It is clear that anthropogenic translocations cause detrimental effects on the native ichthyofauna of the targeted drainage by subjecting the native species to ecological, biological, reproductive, and genetic pressures (Allendorf and Lundquist 2003). Invasive species can eliminate native species genomes via hybridization by genes introgression over generations (Rhymer and Simberloff 1996; Rubidge and Taylor 2004), resulting in the destruction of native species evolutionary heritage (Philipp 1991). From a conservation and management standpoint, this phenomenon is perilous for the redeye bass population.

Besides hybridization, redeye bass population structure is also influenced by geomorphic modification of drainage systems; particularly by streams capture event occurring in headwaters of rivers and streams. Stream capture event is defined as a geomorphologic phenomenon occurring when river drainage system is diverted from its own bed, and flows down the bed of a neighboring stream (Howard and Morgan 1993). Stream capture events between the Savannah

and Santee River drainages, which both harbor native populations of redeye bass, have been recorded (Swift et al. 1986). However, all the natural migrations events are relatively ancient, since that stream capture event was established and dissolved on time scales encompassing geologic eras. With regard to the hypotheses of natural migration, gene flow will occur between ichthyofaunal populations amongst drainages (Ross 1971). Thus, gene flow will eventually lead to the genetic variations between redeye bass populations, which are the focus of study in Chapter 3 of my dissertation.

Dam constructions cause fragmentations of continuous stream ecosystems (Ward and Stanford 1983), which significantly modify flow regimes and hydrologic processes (Mount 1995). Habitat alterations usually favor exotic species, (Godinho and Ferreira 2000; Marchetti and Moyle 2001; Brown and Ford 2002), which further changes the native fish assemblages through predation and competition (Strange et al. 1992; Brown and Moyle 1997). Rehabilitation of the watershed ecosystem is one of the ecologic restoration approach that use fish assemblages with desirable characteristic to replace the altered habitat. This approach requires reconstruction of watersheds with natural flow regime (Poff et al. 1997) and a more sustainable condition (Poff et al. 1997, Fausch et al. 2002). However, deficiency of basic information in setting goals for watershed rehabilitation is often the shortcoming of this approach (Ward et al. 2001).

Microsatellites as the candidate marker

Microsatellites are currently the most widely used genetic markers for fish population level studies (Crooijmans et al.1997). Moreover, the development of microsatellite markers in combination with the improved DNA extraction method from various samples including historical samples also provide a reliable determination of fine-scale population structure as well

as stock composition estimates in mixed-stock fisheries (Nielsen and Hansen 2008). Effective stock identification techniques involve the determination of the origins of fish contributing to the mixed-stock fisheries by comparing the genetic structure of the native samples to the potentially contributing fish populations. There are several main advantages of using microsatellites in population genetic studies which includes their high ubiquity, relative ease of screening once isolated and also the requirement of only a minor amount of lower quality DNA copies for PCR-amplification (Hutchinson et al. 2003). Due to its large number of alleles and high polymorphism, microsatellite variation has also been reported to be powerful in discrimination of weak genetic diversity especially in fish with large populations and high gene flow (Ruzzante et al. 1998; Shaw et al. 1999). Importantly, microsatellites are also co-dominant markers inherited in Mendelian way, which enables direct assessment of genetic variation pattern in various fish population (Salini et al. 2004) as well as aiding the determination of whether the populations are under the Hardy-Weinberg equilibrium or not.

Importance of the population structure study of redeye bass (Micropterus coosae) in Alabama

The alteration of native redeye bass population structure through intra-specific introduction and watershed fragmentation indeed demands attention from both the public and regulatory authorities. This proposed research will be a comprehensive phylogeographic study of all known native redeye bass population within Alabama, which could provide insight as to how external factors may affect the population phylogeny, demographic pattern, and genetic structure. Most importantly, the recovered phylogenies will be used to identify evolutionarily distinct lineages of redeye bass populations.

Redeye bass have also been identified as an important reproductive host for two mussel species, *Hamiota altilis* and *Villosa vibex* (Haag et al. 1999). Freshwater mussel species rely on fish hosts to complete their life history from transition of glochidium (larvae of freshwater bivalves) to the juvenile stage transformation (Zale and Neves 1982; Yeager and Saylor 1995). Both species are endemic to the river drainages where redeye bass are distributed, and *Hamiota altilis* is listed as endangered by the U.S. Fish and Wildlife Service (U.S. Fish and Wildlife Service 1994). Therefore, it is crucial to preserve redeye bass populations because these mussel species require redeye bass as hosts (U.S. Fish and Wildlife Service 1994). Any devastating effects on redeye bass will inadvertently pose threats to the mussel populations as well. Through a future correlative study between mussel species and redeye bass population, a more concise understanding of the interrelation of both species could be developed. Mussel stocking plans will be improved based on their interaction with redeye bass, which is of utmost importance to the conservation of the declining mussel populations.

This study (Chapter 3) should contribute to the planning of biodiversity preservation approaches and implementation of mitigation measures to overcome species extinction problems. Additionally, this study will also provide baseline data upon which sound conservation and management strategies can be formulated. Although restoration and conservation prove to be a huge challenge that requires tremendous effort to achieve desired outcomes, more research in environmental genetics needs to be conducted.With the genetic structure information, detrimental effects which will disrupt the genetic patterns of the native populations could be reduced if not completely eradicated.Thus, the long-term survival of healthy redeye bass populations is the quintessential goal of this work. Ultimately, this work may prove to be a good

start for the long-term preservation, conservation and management of the imperiled freshwater biota of the SE United States.

Research Objectives

The present study focused on the usage of molecular genetic markers in both species and population level studies. DNA barcoding and lab-on-a chip PCR-RFLP methods were utilized for species identification studies while microsatellites were catered for population structure studies. Chapter 1 (DNA barcoding) and Chapter 2 (lab-on-a chip PCR-RFLP) in my dissertation discussed about the usage of these two genetic platforms in catfish species identification. The method of choice for catfish species authentication will be based on the level of resolution and differentiation between different catfish species as well as speed cost of operation, ease of handling, and reproducibility.

Chapter 3 of my dissertation highlighted the study of redeye bass population structure with the following aims: i) Determination of the genetic structure of redeye bass populations among selected water drainages in Alabama by testing the heterogeneity among specimens from distinct geographic areas and ii) Assessment of the level of population differentiation among the redeye bass population in different tributaries of the selected water drainages. The microsatellite dataset were analyzed using several population genetic algorithms and software programs to test the following hypotheses: i) Redeye bass populations from selected drainages will be genetically different; ii) The level of differentiation of redeye bass populations will vary according to different conditions of the selected drainages and their tributaries.

The overall goal of my study was to 1) develop and test methods for catfish species identitification in the context of food safety. Such methods should be useful for regulatory agencies such as United States Food and Drug Administration (FDA) or United States Department of Agriculture Food Safety Inspection Service; and 2) to generate population genetic

information for redeye bass conservation and management useful for various fisheries agencies, particularly those in the Southeastern United States.

CHAPTER 1

DNA BARCODING OF CATFISH: SPECIES AUTHENTICATION AND PHYLOGENETIC ASSESSMENT

Abstract

As the global market for fisheries and aquaculture products expands, mislabeling of these products has become a growing concern in the food safety arena. Molecular species identification techniques hold the potential for rapid, accurate assessment of proper labeling. Here we developed and evaluated DNA barcodes for use in differentiating United States domestic and imported catfish species. First, we sequenced 651 base-pair barcodes from the cytochrome oxidase I (COI) gene from individuals of 9 species (and an Ictalurid hybrid) of domestic and imported catfish in accordance with standard DNA barcoding protocols. These included domestic Ictalurid catfish, and representative imported species from the families of Clariidae and Pangasiidae. Alignment of individual sequences from within a given species revealed highly consistent barcodes (98% similarity on average). These alignments allowed the development and analyses of consensus barcode sequences for each species and comparison with limited sequences in public databases (GenBank and Barcode of Life Data Systems). Validation tests carried out in blinded studies and with commercially purchased catfish samples (both frozen and fresh) revealed the reliability of DNA barcoding for differentiating between these catfish species. The developed protocols and consensus barcodes are valuable resources as increasing market and governmental scrutiny is placed on catfish and other fisheries and aquaculture products labeling in the United States.

Background

Catfish (Order Siluriformes) are a diverse group of fish representing more than 3,000 species, 478 genera and 36 families (Ferraris and de Pinna 1999). Ictalurid catfish represent the largest segment of the domestic aquaculture industry in the United States, generating approximately 600 million pounds of catfish per year (USDA 2002). Imports of Pangasiid, Clariid, and Ictalurid catfish to the United States from East Asia (largely Vietnam and China) have increased rapidly over the last decade and now account for up to half of catfish sales in the U.S. (Hanson 2009). Import restrictions and labeling requirements have impacted the sources and species of imported catfish, but have not substantially reduced import numbers. Anecdotal and documented cases of catfish species mislabeling (either as another catfish species or as a higher value species) are widespread. New regulations currently under development by the federal government will seek to strengthen inspection of domestic and imported catfish, including verification of correct species labeling. Further development and validation of DNA barcoding techniques and consensus sequences for catfish are therefore needed to ensure accuracy in product labeling and informed consumer choices.

DNA barcoding involves the amplification and sequencing of a short universal molecular tag of approximately 650 bp from the 5' region of the mitochondrial cytochrome oxidase I (COI) gene (Hebert et al. 2003; Tavares and Baker 2008). DNA barcoding using COI has been widely employed in various biological fields with proven ability to differentiate closely related species in studies ranging from forensic sciences (Dawnay et al. 2007), molecular systematics (Hardman 2005) to seafood products identification (Botti and Giuffra 2010; Lowenstein et al. 2009; Lowenstein et al. 2010; Steinke et al. 2009). Importantly, community-based efforts to develop extensive DNA barcode libraries, most notably the Barcode of Life Data Systems (BOLD), has

led to the adoption of DNA barcoding technology as the gold standard for species identification and has greatly expanded the power of the technique. The BOLD database provides detailed information of COI-sequenced species including the origin and current location of voucher specimens (Ratnasingham and Hebert 2007). Out of almost 30,000 fish species estimated in the world, barcodes for more than 10,000 fish species are currently recorded in the BOLD database. These COI barcodes are gathered from several sources including the Fish Barcode of Life Initiative (FISH-BOL) (FISH-BOL 2010; Ward et al. 2009) and the Marine Barcode of Life Initiative (MarBOL, http://www.marinebarcoding.org). However, for many species, BOLD barcodes are gleaned from uncurated Genbank records and require additional validation before use.

Here we describe the testing and validation of DNA barcode techniques for domestic and imported catfish including the creation of a DNA barcode database containing eight worldwide commercialized catfish as well as two wild populations of catfish species. Species involved in this study (Table 1) included channel catfish (*Ictalurus punctatus*), blue catfish (*Ictalurus furcatus*) and hybrids of channel (♀) x blue (♂) catfish; Southeast Asian walking catfish (*Clarias batrachus*), broadhead catfish (*Clarias macrocephalus*), basa (*Pangasius bocourti*), swai or sutchi (*Pangasianodon hypopthalmus*); and African sharp-toothed catfish (*Clarias gariepinus*). To complement this study for the purpose of detecting market substitution with vulnerable species, two wild catfish species from China were also included: helmet catfish (*Cranoglanis bouderius*) and long-barbel catfish (*Hemibagrus macropterus*). My results indicate that DNA barcoding is a powerful technique, allowing accurate identification of known, blinded, and commercial samples. As the United States heightens inspection and regulation requirements for

seafood products, DNA barcoding will serve as an important tool in efforts to ensure consumer safety and fair international commerce.

Materials and Methods

Sample Collections

A total of 173 individual samples representing 9 catfish species and an Ictalurid hybrid were used in this study (Table 1). All fin clips or tissue samples were preserved in 95% ethanol (1:10 w: v) upon collection. Ictalurid species were obtained from resource populations of the Department of Fisheries and Allied Aquacultures at Auburn University, USA. Pangasiid and Clariid catfish finclips were obtained from the Department of Aquaculture, Faculty of Fisheries, Kasetsart University, Thailand. In addition, four different types of catfish specimens sold as catfish fillet, catfish nugget and skinless catfish, and swai fillet (swai catfish from Vietnam) were purchased from local grocery stores (Auburn, AL) and oriental markets (Atlanta, GA). Mitochondrial COI sequence data for both *Cranoglanis bouderius* and *Hemibagrus macropterus* were obtained in collaboration with the Laboratory of Fish Phylogenetics, Institute of Hydrobiology, Chinese Academy of Sciences, China.

DNA extraction

Fin clips or muscle tissue samples were used to extract DNA from all samples. Twenty mg starting material was transferred to a 1.5 ml centrifuge tube containing digestion buffer (Liu et al. 1998) and Proteinase K at a concentration of 100 µg/ml. DNA was isolated using the Gentra Puregene Tissue Kit (QIAGEN, USA), following manufacturer's instructions. The

concentration and purity of isolated DNA were estimated using an Ultrospec 1100 Pro spectrophotometer (GE Sciences, NJ, USA) as well as electrophoresis on a 1.5% agarose gel.

PCR Amplification

In order to amplify 651 bp fragment from the 5' end of mitochondrial COI gene, PCR reactions were conducted using primer cocktails of C_FishF1t1 and C_FishR1t1 (Table S1) (Ivanova et al. 2007). The amplification reactions were performed in a total volume of 10 µl and included 1x Invitrogen Platinum Taq Buffer, 0.25 mM each of deoxynucleotide triphosphate (dNTPs), 2.0 mM $MgCl_2$, 10 pmol of each primers, 100 ng of genomic DNA, and 0.5 units of Taq DNA polymerase. The reactions were conducted using a PTC-200 DNA Engine Thermal Cycler (Bio-Rad Laboratories, Hercules, CA) under the following conditions: an initial denaturation at 94 °C for 2 min; 35 cycles of 94 °C for 30 s, 52°C for 40 s and 72°C for 1 min; and concluded with a final elongation step of 72 °C for 10 min followed by a hold at 4 °C (Ivanova et al. 2007). To ensure that the reactions yielded adequate amplicon sizes, PCR products were electrophoresed and visualized on 2.0 % agarose gels containing ethidium bromide (10 mg/ml).

Mitochondrial COI Region Sequencing

Amplified PCR products were subsequently cleaned by the Exo-SAP method (Dugan et al. 2002). Five µl of PCR product, 0.7µl of Exonuclease I 10x Buffer (New England Biolabs, MA, USA), 0.5 µl of Exonuclease I (New England Biolabs Inc., MA, USA), 0.5 µl of rAPid Alkaline Phosphatase (Roche Applied Science, IN, USA), and 5.3 µl of nanopure water were incubated at 37°C for 30 min before being denatured at 80°C for 20 min. The purified products

were labeled using the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, CA, USA) in a total reaction mixture of 10 µl containing 4.94 µl of nanopure water, 1.94 µl of 5x BigDye Buffer (400 mM Tris–HCl pH 9.0 and 10 mM $MgCl_2$), 2 µl of 10 pmol of M13F or M13R (Table S1), 0.12 µl of BigDye Terminator (Applied Biosystems, CA, USA), and 1 µl ExoSAP products. Sequence-PCR products were cleaned up using the ethanol/EDTA precipitation method and sequenced bi-directionally on an ABI 3130 x l Genetic Analyzer (Applied Biosystems, CA, USA). Sequence Analysis Software Version 5.2 (Applied Biosystems, CA, USA) was used to generate sequence tracefiles and contiguous read lengths.

Data Analysis

Sequences were manually assembled using Vector NTI software (Invitrogen, CA, USA). Assembled contigs were end-trimmed to a homologous region using the SeqMan program (DNASTAR, WI, USA). Sequences from vouchered specimens were submitted to the GenBank Barcode database with accession numbers **JF292297-JF292429.** The edited individual contigs for each species were aligned with Vector NTI to produce consensus sequences representing each species. Voucher sequences from GenBank, reference sequences from BOLD databases and consensus sequences of each species generated from this study were compared and aligned using the CLUSTALW program. The multiple sequence alignments were processed using the BOXSHADE 3.21 server (Hoffman and Baron, http://www.ch.embnet.org/software/BOX_form.html) to illustrate the homologous relationship of each species (data not shown). Reference sequence numbers (BOLD) and Accession numbers (GenBank) for voucher species which were used to construct multiple sequence alignments (for *I. punctatus*) are listed in Table S2.

Sample identification based on the sequence similarity approach was carried out using two databases; BOLD and GenBank. The highest percent pairwise identity of the consensus sequence from each species blasted (BLASTN) against NCBI were compared to the percent specimen similarity scores of the consensus sequence from each species within the BOLD-IDS (BOLD Identification System) (Ratnasingham and Hebert 2007). To test the efficiency of DNA barcoding as a species identification tool, a blind sampling test was conducted, in which samples, identity unknown except to the submitting individual, were selected and sequenced.

For sequence comparisons, pairwise genetic distances were quantified based on the Kimura 2-parameter (K2P) distance model (Kimura 1980) using MEGA, version 5.0 (Tamura et al. 2007) A Maximum Parsimony (MP) tree using Close-Neighbor-Interchange algorithm was constructed to display a graphical view of the catfish species studied here (Nei and Kumar 2000). The robustness of the MP tree was assessed by performing bootstrapping analysis with 1000 replicates, and gaps removed by complete deletion (Felsenstein 1985). Confidence levels estimated from the analysis were assigned to each node in the tree and a consensus sequence from *H. macropterus* was used to root the tree.

Results

The mitochondrial cytochrome oxidase I (COI) region of all samples was successfully amplified using PCR. Table 2 shows the comprehensive barcoding identification results based on GenBank and BOLD databases. Both databases revealed definitive identity matches in the range of 96%-100% for consensus sequences of five species (*Ictalurus furcatus*, *Ictalurus punctatus*, *Pangasius bocourti*, *Pangasianodon hypophthalmu*s and *Cranoglanis bouderius*) and an Ictalurid

hybrid. GenBank-based identification for all species yielded an alignment E-value of 0.0. BOLD-IDS results were in agreement with GenBank results in identification of these species, yielding 100% identity, except for *I. furcatus*, *P. bocourti* and *C. bouderius*. For example, *I. furcatus* had 100% maximum identity in Genbank, whereas the percent similarity in BOLD database for this species was 99.41%. Similarly, *P. bocourti* also showed 100% maximum identity in GenBank, whereas the percent similarity for this species in BOLD database was 99.85%.

This study also highlighted, however, existing shortcomings in BOLD and GenBank databases for catfish species. GenBank failed to discriminate *Clarias gariepinus* and *Clarias macrocephalus* from *Clarias batrachus.* At the time of analysis, GenBank only had entries listed as *C. batrachus.* However, the top GenBank hit using our *C. macrocephalus* sequences was a single *C. batrachus* sequence (99% identity). Further investigation and consistent sequences from multiple positively identified *C. macrocephalus* samples led us to conclude that this GenBank *C. batrachus* sequence is mislabeled and truly represents *C. macrocephalus.* Additional *C. batrachus* sequences in GenBank appear also to be mislabeled, and are fairly distantly related to any of the *Clarias* species studied here (87-89% identity). Further identification would be needed to determine whether these sequences represent an isolated branch of *C. batrachus* or, more likely, whether they are truly from another species. Problems with Clariid identification continued in BOLD database. BOLD-IDS rely on GenBank sequences for much of its content and misidentification issues can, therefore, easily be compounded. Our *C. batrachus* sequences returned no match because the BOLD-IDS was relying on GenBank "*C. batrachus*" sequences and uses a 97% identity cutoff in declaring matches. The GenBank *C. batrachus* sequence I had determined represented *C. macrocephalus* was again used by BOLD-

IDS and strongly matched our *C. macrocephalus* sequences (99.69%). BOLD-IDS do include a legitimate *C. gariepinus* barcode and I recorded 99.85% identity matches using our *C. gariepinus* samples. Further, both BOLD-IDS (species level and public data records) and GenBank database were unable to identify *H. macropterus.* No match was garnered for *H. macropterus* from BOLD-IDS, while GenBank, lacking a *H. macropterus* sequence, returned a top hit for a related species, *Hemibagrus velox (*87% identity).

From Table 3, I found that small subsamples of catfish purchased in local grocery and oriental markets were labeled correctly. All the specimens yielded coherent and perfect results (100% matches) in both databases. Interestingly, blue catfish from the USA were more commonly retailed as fresh product in oriental markets than channel catfish, likely indicating a wild-caught fish.

Two specimens from each of the seven species and hybrid catfish (except *C. bouderius* and *H. macropterus*) were randomly selected by a third party for a blind sample test; with the blind sampling test yielding 100% correct species identification results. This result proved that COI barcoding is an efficient tool for unknown species identification with user bias removed.

As shown in Table 4, 651bp of COI consensus barcodes for each species were treated as discrete units to estimate the pairwise level of genetic divergence using the Kimura 2-parameter (K2P) correction model (Nei and Kumar 2000). The K2P distance matrix showed a relatively high overall mean interspecific divergence of 18.3% with a standard error of 1.3%. The K2P distance between species ranged from a low 0.8% (hybrids and *I. punctatus*) to a maximum value of 22.6% (*C. macrocephalus* and Ictalurid hybrid). All the species studied displayed low levels of conspecific divergence.

According to the Maximum Parsimony (MP) tree (Figure1), the species in the present study were clustered independently within their corresponding genera. Three distinct subclades which consist of families Ictaluridae (*Ictalurus*, 2 species and a hybrid), Pangasiidae (*Pangasius* and *Pangasianodon*, 2 species) and Clariidae (*Clarias*, 3 species) were identified; supported by bootstrap values of 99%, 75% and 98% respectively. As presumed, *I. punctatus* and hybrid catfish (*I. punctatus* x *I. furcatus*) formed a cohesive group with a bootstrap value of 100%. Similarly, *C. batrachus* and *C. macrocephalus* created a subclade, which was recognized with a moderately significant boostrap proportion of 0.86. Interestingly, Asian catfish represented by family Pangasiidae did not form an assemblage with another Asian catfish family Clariidae, but was clustered together with family Ictaluridae before merging with Clariidae at a 44% bootstrap value. *H. macropterus* and *C. bouderius* appeared structured as individual subclades away from the other monophlyletic clades.

With the exception of poorly documented or mis-documented catfish species in GenBank and BOLD databases, multiple sequence alignments between consensus sequences (generated from this study) and consensus sequences from the two databases showed high identities (Figure 2). While small variations were observed among fish sequenced within a given species (Table 1), species-specific identifying sequences could be obtained in every case, usually with high concordance with existing database entries. All sequences from vouchered specimens used in the study were submitted to GenBank's Barcode database with accession numbers **JF292297-JF292429**. These sequences are also searchable through cross-referencing in the BOLD database. Additionally, all sequences (including consensus) generated in this study were used to create a searchable database as part of the larger catfish genome database (cBARBEL). The database can be found at http://www.animalgenome.org/catfish/fishid/. Users can search a

barcode of interest again through one or all of the indexed species. The database will be updated as additional sequences and species are added.

Discussion

Species Identification Based on BLAST and BOLD

Regulatory scrutiny of seafood products and their labeling has lagged behind a surge in availability of imported wild-caught and aquaculture species in the United States. The particularly dramatic growth in catfish imports, their impact on the domestic catfish industry, and widespread questions regarding transparency in imported catfish origins and culture conditions, have combined to place catfish at the fore of emerging efforts to heighten fish product inspections in the U.S. A critical component of seafood inspections is determination of accuracy in species labeling. Molecular species identification using DNA barcoding has been applied successfully elsewhere but techniques and consensus barcodes had not been developed and validated in commercial catfish species. In this study, I have sequenced the COI region of the mitochondrial DNA to create a set of barcode sequences used to identify nine catfish species (and an Ictalurid hybrid) from five genera. I extensively compared my results to BOLD and GenBank databases records and found that, out of nine species studied, only five of them matched the reference sequences in both databases. The remaining species that were not perfectly aligned with the two databases included the three Clariid species listed in Table 2 and *H. macropterus*. Both *C. macrocephalus* and *H. macropterus* were yet to be barcoded in BOLD database, whereas *C. gariepinus* lacked any record in the Genbank database. On the other hand, my results brought into question the validity of *C. batrachus* in both databases.

Using our *C. batrachus* sequences as queries against the BOLD-IDS returned "no match." BOLD-IDS validates its identification search only if the species in the reference database has at least three barcoded specimens and identifies the query sequences if it matches the reference sequence within the conspecific distance of less than 2% (Tamura et al. 2007) or not exceeding 3% as suggested by Wong and Hanner (2008). Low (89%) matches were also recorded with *C. batrachus* sequences in GenBank. However, after re-examining the identification and sampling history of these specimens, I strongly suspected that the aberrant results revealed either that the *C. batrachus* sequences stored in both the BOLD and GenBank databases were originally specimens of *C. macrocephalus* or hybridized species that have been accidentally utilized in cataloging the barcodes. Therefore, correct species labeling, morphological taxonomy and voucher documentation should be prioritized in case that reassessment of spurious data is necessary (Ward et al. 2005). Mislabeling is not unexpected since both of these species are genetically homologous (Na-Nakorn et al. 2002) and morphologically similar. It has been reported that *C. macrocephalus* could not be distinguished from a female *C. batrachus* (Teugels et al. 1999). Furthermore, artificial hybridization of *C. macrocephalus* x *C. batrachus* for aquaculture purposes is increasingly popular (Boonbrahm et al. 1977; Smith et al. 2002). Another possible explanation of low or unmatching results for *C. batrachus* specimens is that geographically divergent populations of *C. batrachus* may exist. This has been demonstrated in a karyological study which showed that *C. batrachus* from South Asia is distinctive from populations from Southeast Asia (Garcia-Franco 1993; LeGrande 1981). Therefore, some of the *C. batrachus* specimens' barcoded in both databases may represent a subspecies from South Asia.

I encountered several difficulties in ascertaining the accuracy of BOLD and GenBank records that illustrate current shortcomings in these systems. BOLD data records and sequences often lack transparency for all but the most common species. For example, only one reference sequence for *C. macrocephalus* and C. *batrachus* is available for public viewing, despite more being deposited in the database. Lack of access to these additional sequences makes it hard to ascertain how species determinations are being made using the BOLD database.

Additionally, as mentioned above, a large percentage of publicly available barcodes in BOLD-IDS come from GenBank where there is high probability of tentative, incorrect or low-quality sequences being archived in an era of high-throughput sequencing. Additionally, the accuracy of sequence data cannot be verified given that sequence tracefiles or voucher samples are not retrievable via GenBank. Likewise, difficulties also arise in BOLD database to corroborate suspected records although greater effort is made on quality control (Wong and Hanner 2008). For species with few records, mistakes in private submissions and/or records gleaned from GenBank can result in incorrect identification of samples sequences using the BOLD-IDS. Continual changes to private records and addition and subtraction of sequences can also change identification results obtained over time. Caution and due diligence is therefore required from the user seeking to utilize existing databases for barcode-based species identification.

Sequence Divergence and Phylogenetic Analysis

One crucial barcoding criteria is that congeneric divergence should be higher than conspecific divergence (Hubert et al. 2008). While the sequence variation between five genera observed in this study was atypically high, averaging 18.3 %, other studies showed a lower congeneric variance such as 7.48% in shark and rays (Ward et al. 2008), 8.37% in Canadian freshwater fishes (Hubert et al. 2008), and 9.93% in Australian marine fishes (Ward et al. 2005). In view of this, population genetic and taxonomic analysis will be able to provide a clearer picture of the evolutionary history of catfish in this study. A maximum genetic distance of 3% is sufficient to distinguish all the catfish in this study. As expected, species from the same genera were clustered tightly into a single clade with well supported bootstrap proportion (Steinke et al. 2009). Hybrid catfish with a maternal parent from *I. punctatus*, showed the expected result of barcoding as *I. punctatus* with minimal genetic distance (0.8%) resulting from intraspecific variation within channel catfish.

From Figure1, Pangasiidae was observed as sister group to Ictaluridae albeit at a relatively low bootstrap percentage of 56%, whereas Bagridae represented by *H. macropterus* was the most diverged family from the rest of the groups (Jondeung et al. 2007). Congruent with our data, Funk and Omland (Funk and Omland 2003) has also found that the clustering of *C. macrocephalus* and *C. batrachus* in one lineage and *C. gariepinus* in another lineage resulted from their geographical separation during early stages of their evolution; with the former two species being native Asian catfish and the latter of African origin. The mean genetic distance between these two lineages is 14.7% (Table 4).

In conclusion, DNA barcoding is emerging as an invaluable tool to regulatory agencies and fisheries managers for species authentication, food safety, conservation management as well

as consumer health and support (Costa et al. 2007). Here, I have developed and validated DNA barcoding techniques and consensus sequences for important aquaculture and wild species of catfish. My results indicate that DNA barcoding is a powerful technique, accurately identifying samples regardless of sample source. The barcodes have been deposited in a searchable catfish barcoding database that will be updated as additional samples and species are sequenced. The developed barcodes will aid in upcoming efforts to heighten U.S. fish products inspection and regulation requirements by ensuring accurate labeling of frozen and processed catfish products. Consensus barcodes from these species will also speed the development of fast-turnaround/high-throughput array or SNP-based assays based on informative COI polymorphic sites.

References

1. Boonbrahm M, Tarnchalanukit W and Suraniranat P (1977). Experiments on hybridization of freshwater catfish, *Clarias macrocephalus* Günther and *Clarias batrachus*. Research Report of the Kasetsart University.143p.
2. Botti S and Giuffra E (2010). Oligonucleotide indexing of DNA barcodes: Identification of tuna and other Scombrid species in food products. BMC Biotechnology 10: 60.
3. Costa FO and Carvalho GR (2007). The barcode of life initiative: synopsis and prospective societal impacts of DNA barcoding of fish. Genomics. Society and Policy 3: 29-40.
4. Dawnay NR, Ogden R, McEwing R, Carvalho GR and Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Science International. In press.
5. Donsakul T and Magtoon W (1989). A chromosome study on two species of clariid catfishes, *Clarias batrachus* and *Clarias macrocephalus* from Thailand. Proceedings of the 27th Kasetsart University Annual Conference, Bangkok, Thailand 421-428. Bangkok: Kasetsart University Press.
6. Dugan KA, Lawrence HS, Hares DR, Fisher CL and Budowle B (2002). An improved method for post-PCR purification for mtDNA sequence analysis. Journal of Forensic Science 47: 811-818.
7. Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.
8. Ferraris Jr. CJ and de Pinna MCC (1999). Higher-level names for catfishes (Actinopterygii: Ostariophysi: Siluriformes). Proceedings of the California Academy of Sciences 51: 1-17.
9. FISH-BOL 2010. Fish Barcode of Life Initiative Campaign: Species of Siluriformes yet to be barcoded. Available: http://www.fishbol.org/progress_list_2.php?region=7andlevel=order_nameandtype=Siluriformesandaction=pdfandlist=not. Accessed 2010 Aug 2.
10. Funk DJ and Omland KE (2003). Species-level paraphyly and polyphyly: frequency, causes and consequences, with insights from animal mitonchondria DNA. Annual Review of Ecology, Evolution and Systematics 34: 397-423.
11. Garcia-Franco M (1993) Intraspecific and interspecific relationships of the Clariid catfish *Clarias batrachus*. PhD. thesis. Tokyo University of Fisheries.
12. Hanson T (2009). Catfish database from USDA National Agricultural Statistics Service (NASS) Agricultural Statistics Board, U.S. Department of Agriculture.
13. Hardman M (2005). The phylogenetic relationships among non-diplomystid catfishes as inferred from mitochondrial cytochrome *b* sequences; the search for the Ictalurid sister taxon (Otophysi: Siluriformes). Molecular Phylogenetics and Evolution 37: 700-720.
14. Hebert PDN, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London, Series B 270: 313-321.
15. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, et al. (2008) Identifying Canadian Freshwater Fishes through DNA Barcodes. PLoS ONE 3(6): e2490.
16. Ivanova NV, Zemlak TS, Hanner RH and Hebert PDN (2007). Universal primers cocktails for fish DNA barcoding. Molecular Ecology Notes doi: 10.1111/j.1471-8286.2007.01748.x.

17. Jondeung A, Sangthong P and Zardoya R (2007).The complete mitochondrial DNA sequence of the Mekong giant catfish (*Pangasianodon gigas*), and the phylogenetic relationships among Siluriformes. Gene 387: 49-57.
18. Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 15: 111-120.
19. LeGrande WH (1981). Chromosomal evolution in North American catfishes (Siluriformes: Ictaluridae) with particular emphasis on the madtoms, *Noturus*. Copeia 1981: 33-52.
20. Liu Z, Nichols A, Li P and Dunham RA (1998). Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) and their F_1, F_2 and backcross hybrids. Molecular and General Genetics 258: 260-268.
21. Lowenstein JH, Amato G and Kolokotronis SO (2009). The real maccoyii: Identifying tuna sushi with DNA barcodes-contrasting characteristic attributes and genetic distances. PLoS One 4 (11): e7866.
22. Lowenstein JH, Burger J, Jeitner CW, Amato G, Kolokotronis SO, et al. (2010). DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. Biology Letter 6 (5): 692-5.
23. Na-Nakorn U, Sodsuk P, Wongrat P, Janekitkarn S and Bartley DM (2002). Isozyme variation among four species of the catfish genus *Clarias.* Journal of Fish Biology 60: 1051-1057.
24. Nei M and Kumar S (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.
25. Ratnasingham S and Hebert PD (2007). BOLD: The Barcode of Life Data System. Molecular Ecology Notes 7: 355-364.
26. Smith EG, Leung L, Phaloeun C and Sopheap U (2002). Rice Field Fish Farming Integrated With Rodent Pest Management. In Cambodia. Cambodian Journal of Agriculture.
27. Steinke D, Zemlak TS and Hebert PD (2009). Barcoding nemo: DNA-based identifications for the ornamental fish trade. PLoS One 4 (7): e6300.
28. Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biololgy and Evolution 24: 1596-1599.
29. Tavares ES and Baker AJ (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evolutionary Biology 8: 81.
30. Teugels GG, Diego RC, Pouyaud L and Legendre M (1999). Redescription of *Clarias macrocephalus* (Siluriformes: Clariidae) from Southeast Asia. Cybium 23(3): 285-295.
31. USDA (2002) Catfish Production. National Agricultural Statistics Service (NASS), Agricultural Statistics Board, U.S. Department of Agriculture.
32. Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PD (2005). DNA barcoding Australia's fish species. Philosophical Transaction of Royal Society London B Biological Science 360: 1847-1857.
33. Ward RD, Holmes BH, White WT and Last PR (2008). DNA barcoding Australasian chondrichthyans results and potential uses in conservation. Marine and Freshwater Research 59: 57-71.
34. Ward RD, Hanner R and Hebert PDN (2009). The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329-356.

35. Wong EHK and Hanner RH (2008). DNA barcoding detects market substitution in North American seafood. Food Research International 41: 828-837.

Table 1: Catfish species used in this study

Species name	Common name	Sampling location	Sample size	Similarity within Species (%)
Ictalurus punctatus	Channel catfish	Auburn University, USA	18	98
Ictalurus furcatus	Blue catfish	Auburn University, USA	18	98
I. punctatus x I. furcatus	Hybrid catfish	Auburn University, USA	19	98
Clarias batrachus	Walking catfish	Nakhon Ratchasima Province, NE Thailand	17	97
Clarias gariepinus	African sharp-toothed catfish	Nakhon Ratchasima Province, NE Thailand	19	98
Clarias macrocephalus	Bighead catfish	Faculty of Fisheries, Kasetsart University, Thailand	16	98
Pangasius bocourti	Basa catfish	Yasothon Province, NE Thailand	22	98
Pangasianodon hypophthalmus	Swai or Sutchi catfish	Nakhon Ratchasima Province, NE Thailand	19	98
Cranoglanis bouderius	Helmet catfish	(Guangxi and Guangdong), China	10	94
Hemibagrus macropterus	Long-barbel catfish	(Chongqing, Guangxi, Jiangxi, Sichuan, Hubei, Hunan, Fujian), China	15	98

Sample size : Number of samples

Table 2: Summary of identification based on each species consensus barcoded sequence using BOLD Identification System (BOLD-IDS) and BLASTN search from GenBank

Species studied	BOLD -IDS		GenBank (BLASTN)	
	Species identification	**% similarity**	**Species identification**	**% Max identity**
Ictalurus furcatus	*Ictalurus furcatus*	99.41	*Ictalurus furcatus*	100
Ictalurus punctatus	*Ictalurus punctatus*	100	*Ictalurus punctatus*	100
Hybrid (*I.punctatus* x *I. furcatus)*	*Ictalurus punctatus*	100	*Ictalurus punctatus*	100
Clarias batrachus	**No match***	**0**	*Clarias batrachus*	89
Clarias gariepinus	*Clarias gariepinus*	99.85	***Clarias batrachus****	**87**
Clarias macrocephalus	***Clarias batrachus****	**99.69**	***Clarias batrachus****	**99**
Pangasius bocourti	*Pangasius bocourti*	99.85	*Pangasius bocourti*	100
Pangasius hypophthalmus	*Pangasianodon hypophthalmus*	100	*Pangasianodon hypophthalmus*	100
Hemibagrus macropterus	***No match****	**0**	***Hemibagrus velox****	**87**
Cranoglanis bouderius	Cranoglanis *bouderius*	97.62	Cranoglanis *bouderius*	96

* Asterisk with bolded words corresponds to problematic identifications of species in the present study using either one or both of the databases. Details are further discussed in the text.

Table 3: Description of analyzed local market samples

Species sold as	Country	Consensus identification	% Match
Catfish fillet	USA	*Ictalurus furcatus* (Blue catfish)	100
Frozen Catfish Nugget	USA	*Ictalurus punctatus* (Channel catfish)	100
Skinless catfish	USA	*Ictalurus furcatus* (Blue catfish)	100
Swai fillet	Vietnam	*Pangasianodon hypophthalmus* (Swai catfish)	100

Consensus identification is referred to species identification based on the highest percentage similarity with their corresponding match percentage from both GenBank (BLASTN) pairwise alignment and BOLD-IDS specimen similarity.
Common name of the identified species is written next to the scientific name in parentheses.

Table 4: Estimates of Pairwise Genetic Distances between Catfish Species under Kimura 2-Parameter Model (Kimura 1980).

											Mean Conspecific
	Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Divergence**
1	*I. furcatus*		0.012	0.013	0.022	0.020	0.020	0.019	0.020	0.020	0.020	0.001
2	*I. punctatus*	0.089		0.003	0.020	0.018	0.021	0.018	0.021	0.019	0.020	0.002
3	Hybrid (*I. punctatus* x I. furcatus)	0.096	**0.008***		0.021	0.018	0.021	0.018	0.021	0.020	0.020	0.001
4	*C. batrachus*	0.224	0.205	0.209		0.017	0.016	0.020	0.021	0.020	0.020	0.002
5	*C. gariepinus*	0.220	0.183	0.185	0.148		0.016	0.019	0.020	0.020	0.020	0.007
6	*C. macrocephalus*	0.213	0.220	0.226	0.134	0.146		0.018	0.020	0.019	0.021	0.007
7	*P. bocourti*	0.177	0.181	0.179	0.215	0.193	0.185		0.015	0.020	0.019	0
8	*P. hypophthalmus*	0.177	0.187	0.185	0.221	0.201	0.201	0.116		0.019	0.017	0.003
9	*C. bouderius*	0.185	0.193	0.197	0.200	0.195	0.203	0.201	0.176		0.019	0.009
10	*H. macropterus*	0.201	0.201	0.199	0.216	0.204	0.223	0.185	0.161	0.184		0.016

Pairwise congeneric divergence was denoted by number of base substitutions per site between species (below diagonal) with their corresponding standard error (above diagonal). Complete deletion of all codon position (1st, 2nd, 3rd and noncoding) was employed in this analysis.

*Genetic distance resulting from intraspecific variation between channel catfish (*I. punctatus*) and Ictalurid hybrid catfish (*I. punctatus* x *I. furcatus*). Numbers 1-10 (first horizontal row) correspond to the species in the second column

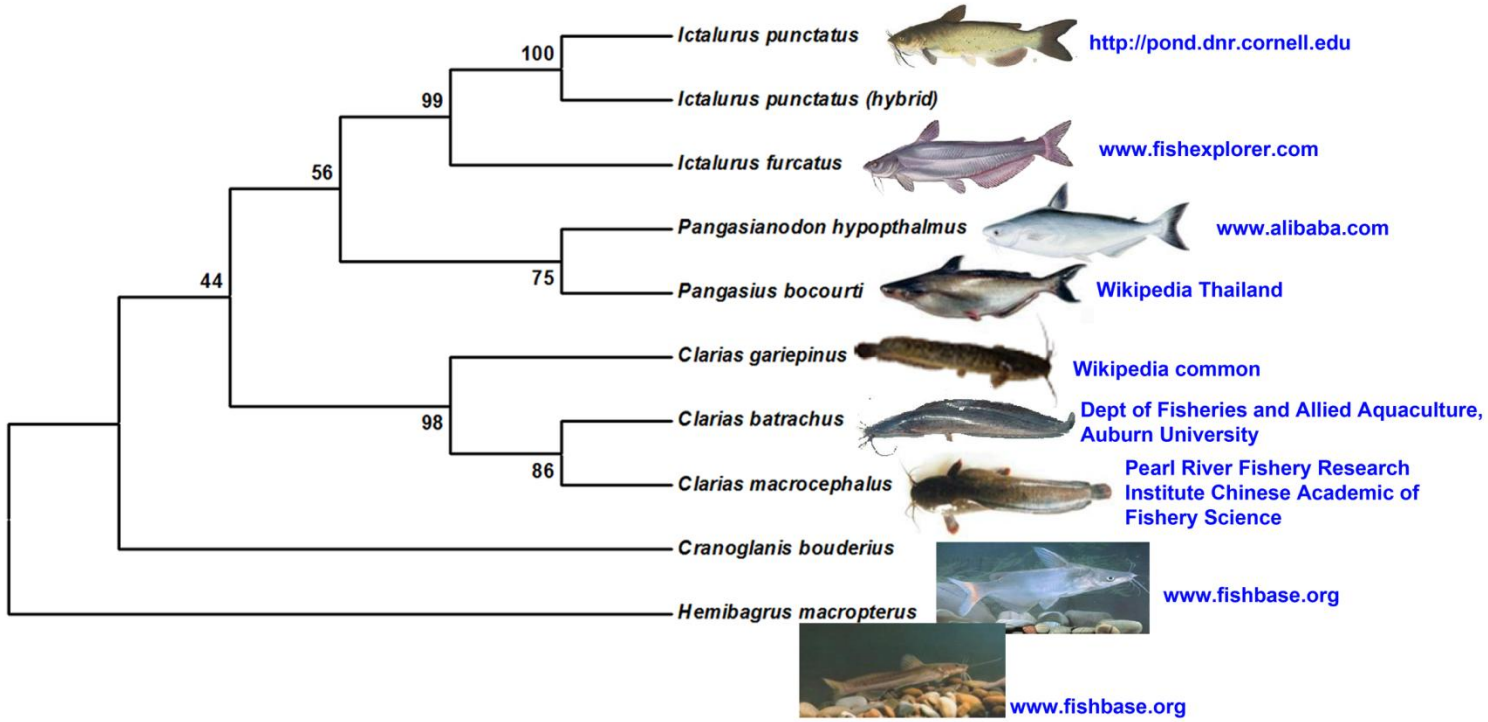
Table S1: Primers used for PCR amplification and sequencing

Name	Primer Sequence (5'-3')	Reference
C_FishF1t1		
VF2_t1	5’TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC3'	Ward et al. 2005
FishF2_t1	5’TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC3'	Ward et al. 2005
C_FishR1t1		
FishR2_t1	5’CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA3'	Ward et al. 2005
FR1d_t1	5’CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA3’	Ivanova et al. 2007
M13F	5’TGTAAAACGACGGCCAGT3'	Messing 1983
M13R	5’CAGGAAACAGCTATGAC3’	Messing 1983

Table S2 Reference sequence numbers (BOLD) and accession numbers (GenBank) of voucher species used to build multiple sequence alignment of *Ictalurus punctatus* using CLUSTALW program in Figure 2.

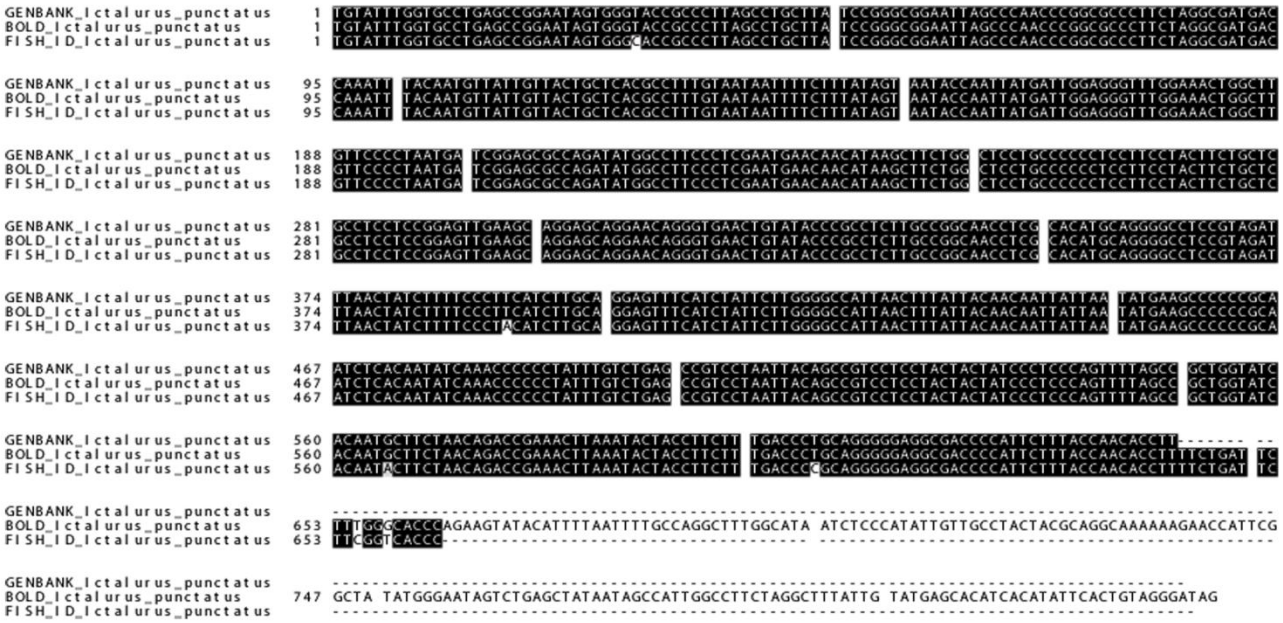
No	BOLD Reference Sequences:	GenBank Accession Number:
1	GBGC0208-06|AF482987	EU752103.1
2	GBGC1573-06|NC_003489	EU752102.1
3	GBGC3968-07|AF482987 AF032380 AF032381	EU752101.1
4	GBGC6278-08|AF482987 AF032380 AF032381	EU752100.1
5	BCFB138-06|BCF-0501-1	EU490865.1
6	BCFB139-06|BCF-0501-2	EU524686.1
7	BCFB129-06|BCF-0114-1	EU524684.1
8	BCFB130-06|BCF-0114-2	EU524683.1
9	BCFB131-06|BCF-0114-3	EU524682.1
10	NHFEC062-06|BCF-0113-1	EU524681.1
11	BCFB132-06|BCF-0115-1	EU524680.1
12	BCFB133-06|BCF-0115-2	EU524679.1
13	BCFB134-06|BCF-0115-3	EU524678.1
14	BCFB135-06|BCF-0394-1	EU524677.1
15	BCFB136-06|BCF-0394-2	EU524676.1
16	BCFB137-06|BCF-0394-3	EU524106.1
17	GBGC3998-08|EU490865	
18	GBGC8082-09|AF482987 AF032380 AF032381	

Figure 1: Phylogenetic consensus tree of nine catfish species (and an Ictalurid hybrid) constructed using Maximum Parsimony (MP) Method.



The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein 1985). *Hemibagrus macropterus* was used as an outgroup. The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Nei and Kumar 2000) with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The source for each image was displayed next to the pictures.

Figure 2: Multiple sequence alignment of consensus sequences for *Ictalurus punctatus* against voucher and reference sequences from GenBank and BOLD databases.



A multiple sequence alignment of *Ictalurus punctatus* was generated by ClustalW and graphically represented by BOXSHADE 3.21. The nomenclature of the aligned sequences is as follows: FISH_ID_Ictalurus_punctatus (as *I. punctatus* consensus sequence in the present study), GENBANK_Ictalurus_punctatus (as GenBank voucher species consensus sequence) and BOLD_Ictalurus_punctatus (as BOLD species reference consensus sequence). Both voucher and reference sequences were downloaded from the two databases with the accession numbers listed in Table S2. Highly conserved regions which were ≥ 50% identical were boxed in solid black and light shading indicates conservative substitutions.

CHAPTER 2

CATFISH SPECIES IDENTIFICATION USING LAB-ON-CHIP POLYMERASE CHAIN REACTION-RESTRICTION FRAGMENT LENGTH POLYMORPHISM (PCR-RFLP)

Abstract

Lab-on-a-chip based Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR–RFLP) technology was developed for the identification of seven catfish species including an Ictalurid hybrid. RFLP profiling of mitochondrial cytochrome b fragments digested by three different restriction endonucleases (*Dde*I, *Hae*III and *Nla*III) was constructed from morphologically verified catfish samples on an Agilent 2100 Bioanalyzer. High sensitivity and ease of handling of this endpoint microchip-based capillary electrophoresis increased resolution and accuracy of DNA fragments sizing. RFLP profiles derived from a combination of all three enzymes produced consistent species-specific identification profiles. Observed restriction fragment patterns for *Clarias batrachus* and *Ictalurus punctatus* were almost similar but both of these species could be differentiated using a single band of *Hae*III restriction site. Advantages due to shorter assay times, assay ease, and minimum usage of harmful solvents and chemicals when compared to traditional DNA barcoding are counterbalanced by the need to develop and optimize specific restriction digest profiles for all potential species of interest. With further development this method may be utilized in catfish species scrutiny to ensure the enforcement of seafood labeling regulations.

Background

Catfish have been extensively marketed in both fresh and frozen forms and constitute the primary farm-raised fish in the United States. A rising market share for imported catfish has introduced mislabeling and fraudulent substitution of higher value species for others of lower-price, as seen previously in other seafood sectors. Mislabeling has taken the form of substitution of Asian basa and tra catfish as Ictalurid catfish or fraudulent country of origin labeling, impacting the profitability of the US catfish industry as well as raising public health concerns (Brambilla et al. 2009). Marketing of fish as processed fillet and seafood products has exacerbated the difficulties of species authentication because distinct morphological characteristics of the fish are removed during processing stage (McDowell and Graves 2002). Hence, the development of innovative, rapid and low cost analytical platforms for species identification is of high importance to guarantee the authenticity and origin traceability of catfish along the seafood distribution chain from pond to plate.

Numerous protein-based analyses have been developed specifically for species authentication (Tepedino et al. 2001; Ochiai et al. 2001). However, these methods are only reliable for identification of raw fish and not applicable for processed fish products as the thermally labile proteins become irreversibly denatured by heat (Dooley et al. 2005). Alternatively, DNA-based techniques are a more efficient identification approach, largely because of their high thermal stability as well as their independence of cell types and age (Davidson 1998; Bossier 1999; Lockley and Bardsley 2000).

Although DNA-based methods such as real-time PCR and direct sequencing provide definitive species identification, their higher cost or longer assay times, respectively, are

drawbacks for their use as a robust and routine fish screening test (Meyer et al.1995). Therefore, a relatively simple and inexpensive species discrimination technique such as PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) may be more appropriate for targeted, rapid species identification. In fact, PCR-RFLP mapping has been successfully applied in various fish species and fish products identification (Arahishi 2005; Akasaki et al. 2006; Lin and Hwang 2007; Ce´spedes et al. 1998; Hsieh et al. 2007). Briefly, this technique incorporates the analysis of specific DNA fragment profiles generated by digestion of PCR amplicons with selected restriction endonucleases to produce restriction fragment patterns through electrophoretic sorting. However, PCR-RFLP still relies on conventional gel electrophoresis for endpoint detection that is technically hazardous, time demanding and complicated by the production of variable results (Dooley et al. 2005). The complex DNA fingerprints and sizing variation has led to the development of micro-chip based capillary electrophoresis technology such as automated lab-on-chip electrophoretic system replacing the traditional gel electrophoresis step (Fajardo et al. 2009; Panaro et al. 2000). The Agilent 2100 Bioanalyzer is the first commercialized chip-based electrophoresis apparatus that utilize capillary electrophoresis (CE) technology. This automated system involves DNA fragments separation by CE in miniaturized channels of the microfluidic chip followed by detection steps using laser-induced fluorescence (Vasilyeva et al. 2004). This user-friendly technology offers vast advantages of speed and safety, permitting the generation of higher resolution PCR-RFLP profiles in comparison to conventional gel-based approaches (Uthayakumaran et al. 2005). Though the overall cost of this advanced device is fairly high, it is still ideally suitable for routine surveys compared to the costly real-time PCR method (Dooley and Garrett 2001).

With the collaboration of Agilent Technologies, I developed PCR-RFLP profiles on the 2100 Bioanalyzer based on the mitochondrial cytochrome b gene, to enable identification of seven catfish species (2 Ictalurid species and an Ictalurid hybrid; 3 Clariid species; and 2 Pangasiid species) that are commercially important in the seafood industry. Sequences generated from the cytochrome b gene were also used as supporting analysis to confirm the competence of the PCR-RFLP technique. Overall, my results demonstrated that the Lab-on-chip technique may be rapidly performed for catfish species authentication by regulatory agencies to enforce transnational laws and regulations, protecting the consumer against seafood fraud.

Materials and Methods

Sample Collections

At least 12 individuals from each catfish species were morphologically identified by a fish taxonomist at the respective sources before the samples were obtained. All fin clips or tissue samples were preserved in 95% ethanol (1:10 w: v) upon collection. Details of specimens are presented in Table 5.

DNA extraction

Total genomic DNA was isolated from fin clips or muscle tissues. In brief, 20 mg of samples were incubated overnight in digestion buffer (Liu et al. 1998) containing 100 ug/ ml Proteinase K at 55°C. The subsequent extraction process was based on a slight modification of the Gentra Puregene Tissue Kit protocol (Qiagen, Valencia, CA). Final concentration and purity of extracted DNA was estimated by means of Ultrospec 1100 Pro spectrophotometer (GE

Sciences, Pistachaway, NJ). DNA quality of the samples was examined by electrophoresis on a 1.5% agarose gel containing ethidium bromide.

PCR Amplification

Amplification of the partial mitochondrial cytochrome *b* gene was carried out in a total volume of 25µl containing 50 ng of the DNA template and components from the Fish Species ID Beta Kit (Agilent Technologies, Santa Clara, CA) which consist of sterile water, 2 x PCR Master Mix and a primer mix. The primer mix is constituted of the following pair of universal primers: L14735 (5'-AAA AAC CAC CGT TGT TAT TCA ACT A-3') and H15149ad (5'-GCI CCT CAR AAT GAY ATT TGT CCT CA-3') which were described by Russell et al. (2000) and Wolf et al. (2000) respectively. PCR mixture was performed according to the kit's manual and the reaction was carried out in a PTC-200 DNA Engine Thermal Cycler (Bio-Rad Laboratories, Hercules, CA) programmed as follows: 5 min at 95°C, 40 cycles of 30 s at 95°C, 30 s at 50°C and 30 s at 72°C, and a final elongation step at 72 °C for 7 min. For PCR quality assurance, each assay was accompanied by a negative control (no-template blank) and a positive control (*Oncorhynchus keta* reference specimen). PCR products were examined using the DNA1000 LabChip to confirm amplification.

Restriction Digestion

Unpurified PCR amplicons were subjected to three independent restriction reactions with *Dde*I, *Hae*III and *Nla*III enzymes. The amplified fragments were digested for at least 2 hours to overnight with 1 unit of each restriction endonuclease in a total volume of 5µl at 37°C and the reactions were inactivated by incubation at 65°C for 15 min.

PCR-RFLP Profiling

PCR-RFLP profiles were visualized using DNA1000 LabChips with the 2100 Bioanalyzer microchip capillary electrophoresis system (Agilent Technologies, Santa Clara, CA). Both DNA1000 LabChips and reagents were prepared as required according to the manufacturer's instructions. Digested PCR products were mixed with 60mM EDTA before being loaded on the microchip. RFLP fingerprints were obtained within 30 min and were analyzed using 2100 Expert software and compared to the restriction map generated from the sequence analysis.

Mitochondrial Cytochrome b Gene Sequencing

To confirm restriction fragment patterns, PCR products were sequenced bi-directionally with the same primers used for PCR reaction. PCR amplicons were cleaned up before sequencing reaction following the protocol of Exo-SAP method as described in Dugan et al. 2002. Direct sequencing was accomplished on an ABI 3130 x l Genetic Analyzer using the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Sequence Analysis Software version 5.2 (Applied Biosystems, Foster City, CA) was used to generate sequence tracefiles and contiguous read lengths.

Sequence Analysis

Sequences were end-trimmed using the SeqMan program (DNASTAR, Madison, WI). Based on the sequence similarity approach, specimen identification was also carried out by comparing blasting the sequences against the GenBank database. Multiple sequence alignment was performed to create consensus sequence for each catfish species using the CLUSTALW

program (Thompson et al. 1994). A phylogenetic tree was constructed from the species consensus sequence by the maximum parsimony (MP) algorithm (Nei and Kumar, 2000) along with bootstrapping analysis using MEGA 5.0 (Tamura et al. 2007). A distantly related catfish *Cranoglanis bouderius* was used as an out-group for this MP tree. In addition, Neighbor-Joining (NJ) and Maximum Likelihood (ML) methods were also performed to confirm the absence of clustering output variation between different phylogenetic reconstruction methods.

These sequences were also used to produce predicted RLFP patterns generated by the three enzymes. These restriction maps were constructed using the software NEBcutter (Vincze et al. 2003). A species identification flowchart was built to simplify the analysis process (Figure 5). All observed fragments sizes were rounded to account for fragment size variation.

Results and Discussion

In order to detect commercial fraud in marketing of catfish products, the development of a reliable, rapid traceability technique is necessary. Therefore, an extension of the PCR amplification method using RFLP analysis that has previously proven successful for fish species authentication was used (Asensio Gil 2007; Mackie et al. 1999). To improve the detection and resolution of the RFLP patterns, a highly sensitive, rapid (less than one working day) and feasible end-point detection method based on a capillary electrophoresis lab-on-chip device (2100 Bioanalyzer) was utilized to identify seven catfish species and an Ictalurid hybrid species. A complimentary sequence analysis on the cytochrome b region for each species studied was also conducted to compare the generated RFLP profiles. Catfish species were also confirmed by

comparing the sequences against all entries in GenBank database (Table 7). No ambiguities were present in the alignment of all the catfish species sequences.

The experimental restriction profile revealed that some catfish species identification could be achieved using only a single endonuclease. Based on Table 6, *Pangasius bocourti* displayed a unique restriction RFLP pattern when digested with *Nla*III. Likewise, no other catfish species produced fragment profiles similar to that of *Ictalurus furcatus* when restricted with enzyme *Hae*III. However, a combination of three enzymes was required to produce species-specific RFLP profiles in order to distinguish the other five catfish species in the present study. For instance, the three Clariid species generated almost identical profiles from two enzymes, *Hae*III and *Nla*III, but were only differentiated following the restriction by *Dde*I. Digestion using the *Dde*I restriction enzyme yielded two fragments at size ranges of 196-208 bp and 261-275 bp from the PCR amplicons of *Clarias batrachus*, whereas PCR fragment of *Clarias gariepinus* were not cleaved due to the absence of the *Dde*I restriction site. Despite nearly identical patterns between *Pangasius bocourti* and *Pangasianodon hypopthalmus*, species identification was still possible due to comparison of the restriction patterns of *Hae*III. Two fragments were detected in *Hae*III site of *P. bocourti*, while amplification fragments of *P. hypopthalmus* remained undigested. Therefore, the analysis of all three enzymes restriction patterns were able to resolve all the catfish species under study. Similarly, *C. batrachus* and *Ictalurus punctatus*, which shared almost similar restriction profiles, displayed slight differences in the *Hae*III (Table 6) which enable them to be differentiated. As expected, both *I. punctatus* and the Ictalurid hybrid generated identical RFLP patterns due the fact that Ictalurid hybrid inherited the mtDNA cytochrome b gene from the maternal parent channel catfish (*I. punctatus*).

Predictions based on *in silico* restriction of generated sequences showed that *C. macrocephalus* produced DNA fragments of 4, 6, 124-127, 290-294 bp when digested with the *Hae*III enzyme. However, PCR-RFLP experimental results revealed only two cleaved fragments (Table 6). Similarly, the absence of these theoretically shorter fragments of ~25 bp was also observed in both *I. punctatus* and *I. furcatus* following restriction with the *Nla*III enzyme. This was probably because these small fragments were too weak to be detected by the Bioanalyzer 2100 due to poor fluorescence of these small and low concentration fragments. Fragments smaller than 40 bp were also inconsistently detected. These observations were concordant with the other documentation indicating the difficulties of Bioanalyzer 2100 in resolving fragments which are close to the minimum sizing limits of approximately 25 bp (Dooley et al 2005; Ogden and McEwing 2008). Although these fragments were not observed as expected, the ability to identify catfish species were not affected.

The absence of a 70 bp band following *Nla*III digestion of *P. bocourti* was clearly problematic and possibly related to the flagged warning of "possible co-migration of 2 peaks" indicated by the 2100 Expert software. The loss of this fragment predicted in the *in silico* restriction pattern was possibly explained by the co-migration of larger fragments as a single band (Table 6). In contrast, the missing DNA fragments of approximately 162-166 bp following *Dde*I restriction of *C. macrocephalus* could not be interpreted clearly since there was no evidence of co-migration effect on these sites. Instead, the fragment patterns of approximately 65-69 and 198-207 bp were consistently present in all the individuals of *C. macrocephalus* digested with *Dde*I, indicating that this species could still be differentiated unambiguously.

The RFLP profiles were observed to be largely consistent in all the tested individuals for each catfish species in this study. In fact, the majority of the experimental RFLP patterns

matched those that were predicted from simulated restriction using NEB cutter, except for the absent smaller (<25 bp) fragments. However, variation in the fragment sizes resolved by 2100 Bioanalyzer was observed. Observed restriction patterns were often larger by 10-40 bp than expected based on *in silico* predictions. Indeed, in some cases >10% variation was observed, especially in profiles with smaller fragment sizes, where both absolute error and percentage error were large (Ogden and McEwing 2008). This discrepancy was strongly apparent in fragments of Clariid species digested by *Hae*III as well as *Dde*I-digested fragments of Pangasiid species (Table 6), with one of the fragment sizes consistently observed larger than those expected by approximately 38 bp. For example, in *C. batrachus*, the expected fragment size was 292-293 bp while corresponding observed fragments were 325-336 bp. It was initially assumed that these discrepancies were caused by technical errors in either sequencing or restriction analysis. However, after thorough comparison with the species-specific sequence entries from GenBank, a similar pattern was also seen for the *in silico* restriction of reference sequences for these Clariid and Pangasiid species. Reconfirmation through replication of PCR-RFLP and 2100 Bioanalyzer analysis has suggested that these larger fragments were not artifacts as the DNA sizing results remained consistent, returning the same profile for each species in each replicate. These divergences may be explained by the level of variation among chips and machines which have been reported in earlier studies (Ogden and McEwing, 2008) or differences in digestion efficiency. Reproducibility testing of the 2100 Bioanalyzer results in catfish species identification conducted by other laboratories may be required. Despite these discrepancies, species-specific banding patterns for both expected and observed results were relatively concordant, permitting their reliable identification.

Overall results obtained from the restriction analysis of the mtDNA cytochrome b gene using three enzymes *Dde*I, *Hae*III and *Nla*III were displayed in the computer-generated gel image in the 2100 Expert software (Figure 3). The RFLP fingerprints generated by these three enzymes complemented each other in the catfish species identification. Additionally, the species identification flowchart also serves as a supporting tool for gel image interpretation. The single band observed in the fragments of *C. gariepinus* digested by *Dde*I (Figure 3a) indicated the absence of a *Dde*I restriction site in this particular species. On the other hand, accurate identification of other species needs to be facilitated by the electropherograms corresponding to each band (Figure 4) due to the slight variations in band intensity displayed in the gel image.

A species identification flowchart (Figure 5) was developed as a rapid means for catfish species identification following the development of PCR-RFLP profiles on the 2100 Bioanalyzer. Species diagnosis could be made by comparing the experiment banding patterns to the established reference profiles within the flowchart in a few basic steps which focus on the comparison between the banding patterns in the 2100 Bioanalyzer output and the key description in the flowchart.

To ensure the reliability of the PCR-RFLP technique, initial sequence analysis needs to be performed because PCR-RFLP profiles alone lack the ability to distinguish among other species not examined here. Therefore, all the vouchered catfish specimens used in this study were sequenced. Sequences from at least five individuals were aligned to produce species-specific consensus sequences before the consensus sequences were compared against the GenBank database (Table 7). All the samples revealed relatively high identity matches with GenBank entries in the range of 96-100%. Although there were variations of fragment sizing between the expected restrictions profiles based on the *in silico* restriction analysis of sequences

and the observed profiles generated by the 2100 Bioanalyzer, their restriction fragment patterns were still in concordance with each other. In fact, the cytochrome b gene sequences were aligned well within each species with an average alignment score of 96%. A potential cause of sequence and fragment discrepancies may be the redundant bases within the universal primers which were unable to provide exact match with any specimens studied. This observation was more profound in the sequence fragments of five Asian species (Clariid and Pangasiid catfish) which were consistently shorter than the observed restricted fragments. In fact, the incorporation of an artificial *Dde*I site when a "G" base is present, or disruption of *Dde*I site when an "A" is present, have been previously reported to create doublet in PCR products (Dooley et al. 2005).

A phylogenetic tree was constructed from consensus sequences for each species to verify the presence of sufficient sequence variation between catfish species studied. Phylogenetic trees were obtained from Maximum Parsimony (Figure 6) and Neighbor-Joining and Maximum Likelihood analysis (not shown) produced similar clustering patterns. Tree topologies generated by the present study was similar to those of a previous study using cytochrome oxidase I (COI) (Wong et al. 2011) where the catfish species were grouped into three distinct clades of families Ictaluridae, Pangasiidae and Clariidae. Family Pangasiidae was more closely related to Ictalurid species compared to Clariid species, although the latter family was derived from the same geographical region as Pangasiid species. On the other hand, hybrid catfish with a maternal parent from *I. punctatus*, showed the expected result by forming a cohesive subclade with *I. punctatus*. Considering the intra-specific variations that can influence the reliability of PCR-RFLP analysis, all the replicated specimens were collected from different sampling locations.

Conclusion

My results from the PCR-RFLP of the mitochondrial cytochrome b gene were able to discriminate all the studied catfish species. The Agilent 2100 Bioanalyzer is designed to allow rapid sizing of the digested fragments. By means of DNA1000 LabChips, simultaneous post-digest analysis of 4 samples can be tested in a single run within 30 minutes, which could considerably reduce the amount of time needed for rapid small-scale species identification compared to direct sequencing. The results obtained in this study show that the usage of this system for catfish species identification is potentially feasible and offers several advantages over conventional gel-based electrophoresis. One of the advantages is the improved resolution of the RFLP patterns, through the detection of smaller fragments that are undetectable on conventional gels. Furthermore, the small-sized LabChips are also easier to be manipulated, requiring only a small amount of operation space. In addition, overall PCR and restriction enzyme costs could also be significantly reduced, due to the small volumes that are needed. Compared to conventional gels, this microchip-based system also allows reduction in the usage of harmful DNA staining chemicals used for detection of the RLFP fingerprints. In conclusion, this analytical platform offers several advantages over the standard gel electrophoresis, including speed, ease of handling, higher resolution and minimum usage of harmful solvents and chemicals (Doodley et al. 2005; Spaniolas et al. 2006). Owing to these advantages, this method may potentially be routinely adopted for the authentication of catfish species and thus aid in the enforcement of labeling regulations to ensure safety and quality of fish and their value-added products.

References

1. Akasaki, T., Yanagimoto, T., Yamakami, K., Tomonaga, H., and Sato, S. (2006). Species Identification and PCR-RFLP Analysis of Cytochrome b Gene in Cod Fish (Order Gadiformes) Products. Journal of Food Science 71 (3): 190-195.
2. Arahishi, F. (2005). PCR-RFLP analysis of nuclear nontranscribed spacer for mackerel species identification. Journal of Agricultural and Food Chemistry 53:508-511.
3. Asensio Gil, L. (2007). PCR-based methods for fish and fishery products authentication. Trends in Food Science and Technology 18: 558-566.
4. Bossier, P. (1999). Authentication of seafood products by DNA patterns. J Food Sci 64: 189-193.
5. Cespedes, A., Garcia, T., Carrera, E., Gonzalez, I., Sanz, B., Hernandez, P. E., and Martin, R. (1998). Identification of flatfish species using polymerase chain reaction (PCR) amplification and restriction analysis of the cytochrome b gene. Journal of Food Science 63: 206-209.
6. Davidson, W. S. (1998). PCR techniques. In P.R. Ashurst and M.J. Dennis (Eds.), *Analytical methods of food authentication* (pp. 182-203). London: Blackie Academic and Professional.
7. Dooley, J. J., and Garrett, S. D. (2001). Development of meat speciation assays using the Agilent 2100 Bioanalyzer. Agilent technologies application note no. 5988-4069EN.
8. Dooley, J. J., Sage, H. D., Brown, H. M., and Garrett, S. D. (2005). Improved fish species identification by use of lab-on-a-chip technology. Food Control 16: 601-607.
9. Dugan, K.A., Lawrence, H.S, Hares, D.R., Fisher, C.L. and Budowle, B. (2002). An improved method for post-PCR purification for mtDNA sequence analysis. J Forensic Sci 47: 811-818.
10. Fajardo, V., Gonzalez, I., Dooley, J., Garret, S., Brown, H. M., Garcia, T., et al. (2009). Application of polymerase chain reaction–restriction fragment length polymorphism analysis and lab-on-a-chip capillary electrophoresis for the specific identification of game and domestic meats. Journal of the Science of Food and Agriculture 89: 843-847.
11. Hsieh, H. S., Chai, T., and Hwang, D. F. (2007). Using the PCReRFLP method to identify the species of different processed products of billfish meats. Food Control 18: 369-374.
12. Lin, W. F., and Hwang, D. F. (2007). Application of PCR-RFLP analysis on species identification of canned tuna. Food Control 18: 1050-1057.
13. Liu, Z., Nichols, A., Li, P., and Dunham, R.A. (1998). Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) and their F_1, F_2 and backcross hybrids. Molecular and General Genetics 258: 260-268.
14. Lockley, A.K., Bardsley, R.G. (2000). DNA-based methods for food authentication. Trends Food Sci Technol 11: 67-77.
15. Mackie, I. M., Pryde, S. E., Gonzales-Sotelo, C., Medina, I., Pérez-Martín, R., Rey-Mendez, M., Quinteiro, J., and Rehbein, H. (1999). Challenges in the identification of species of canned fish. Trends in Food Science and Technology, 10, 9-14.
16. McDowell, J., and Graves, J. E. (2002). Nuclear and mitochondrial DNA markers for specific identification of istiophorid and xiphiid billfishes. Fishery Bulletin, 100, 537-544.
17. Meyer, R., Ho¨ felein, C., Lu¨ thy, J., and Candrian, U. (1995). Polymerase chain reaction-restriction fragment length polymorphism analysis: a simple method for species identification in food. Journal of the AOAC International 78: 1542-1551.
18. Nei, M., and Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

19. Ochiai, Y., Ochiai, L., Hashimoto, K., and Watabe, S. (2001). Quantitative estimation of dark muscle content in the mackerel meat paste and its products using antisera against myosin light chains. Journal of Food Science 66: 1301-1305.
20. Ogden, R. and McEwing, R. (2008). Adaptation of DNA analysis techniques for the identification of illegally imported bushmeat for use on the Agilent 2100 bioanalyser. Final Technical Report for Food Standards Agency United Kingdom Project Q01109: 1-20.
21. Panaro, N. J., Yuen, P. K., Sakazume, T., Fortina, P., Kricka, L. J., and Wilding, P. (2000). Evaluation of DNA fragments sizing and quantification by the Agilent 2100 Bioanalyzer. Clinical Chemistry 46: 1851-1853.
22. Russell, V. J., Hold, G. L., Pryde, S. E., Rehbein, H., Quinteiro, J., Rey-Mendez, M., Sotelo, C. G., Perez- Martin, R. I., Santos, A. T., and Rosa, C. (2000). Use of restriction fragment length polymorphism to distinguish between salmon species. J. Agric. Food Chem. 48: 2184-2188.
23. Spaniolas, S., May, S. T., Bennett, M. J., and Tucker, G. A. (2006). Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. J. Agric. Food Chem. 54: 7466-7470.
24. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.
25. Thompson, J. D, Higgins, D. G, Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
26. Tepedino, V., Berrini, A., Borromeo, V., Gaggioli, D., Cantoni, C., Manzoni, P., and Secchi, C. (2001). Identification of commercial fish species belonging to the orders pleuronectiformes and gadiformes: library of isoelectric focusing patterns. Journal of AOAC International, 84(5):1600-1607.
27. Uthayakumaran, S., Batey, I. L. and Wrigley, C.W. (2005). On-the-spot identification of grain variety and wheat-quality type by Lab-on-a-chip capillary electrophoresis. Journal of Cereal Science 41: 371-374.
28. Vasilyeva, E., Woodard, J., Taylor, F.R., Kretschmer, M., Fajardo, H., Lyubarskaya, Y., Kobayashi, K., Dingley, A., Mhatre, R. (2004). Development of a chip-based capillary gel electrophoresis method for quantification of a half-antibody in immunoglobulin G4 samples. Electrophoresis 25(21-22): 3890-3896.
29. Vincze, T., Posfai, J., Roberts, R.J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 31: 3688-3691.
30. Wolf, C., Burgener, M., Hubner, P., and Luthy, J. (2000). PCR-RFLP analysis of mitochondrial DNA: Differentiation of fish species. Lebensm. Wiss. Technol. 33: 144-150.
31. Wong, L.L., Peatman, E., Lu, J., Kucuktas, H., He, S., Zhou, C., Na-nakorn, U. and Liu, Z. (2011). DNA Barcoding of Catfish: Species Authentication and Phylogenetic Assessment. PLoS ONE 6(3): e17812. doi:10.1371/journal.pone.0017812

Table 5: Catfish species used in the development of PCR-RLP profiles

Species name	Common name	Sampling location	Sample size
Ictalurus punctatus	Channel catfish	Auburn University, USA	21
Ictalurus furcatus	Blue catfish	Auburn University, USA	18
I. punctatus x I. furcatus	Hybrid catfish	Auburn University, USA	12
Clarias batrachus	Walking catfish	Nakhon Ratchasima Province, NE Thailand	15
Clarias gariepinus	African sharp-toothed catfish	Nakhon Ratchasima Province, NE Thailand	14
Clarias macrocephalus	Bighead catfish	Faculty of Fisheries, Kasetsart University, Thailand	12
Pangasius bocourti	Basa catfish	Yasothon Province, NE Thailand	14
Pangasianodon hypophthalmus	Swai or Sutchi catfish	Nakhon Ratchasima Province, NE Thailand	15

Sample size: Number of samples collected

Table 6: Predicted and observed PCR-RFLP fragment sizes generated with three different restriction endonucleases *Dde*I, *Hae*III, and *Nla*III of catfish species used in this study. E: Predicted fragment sizes were generated from *in silico* restriction analysis of cytochrome b sequences. O: Observed sizes are the mean sizes obtained from analysis performed on at least 12 catfish individuals in each species. * indicates small fragments which are not detected by 2100 Bioanalyzer while bolded values indicates observed fragments size which are larger than the predicted fragments size.

Species		DdeI	HaeIII	NlaIII
Clarias batrachus	**E**	164-165 266-272	138-144 292-293	30-36 85-86 315
	O	196-208 261-275	140-144 **325-336**	37-38 80-86 311-336
Clarias gariepinus	**E**	424-445	132-137 286-294	26-29 79-87 315
	O	460-484	141-143 **328-337**	37-39 84-87 322-333
Clarias macrocephalus	**E**	60 162-166 200-202	4 6 124-127 290-294	83-87 340-341
	O	65-69 * 198-207	* * 131-138 **320-335**	35-39 81-86 317-365
Pangasius bocourti	**E**	158-162 248-273	123-128 292-297	23-26 70 79-84 91 154
	O	**199-208** 265-274	133-138 332-344	37-39 * 82-86 95-102 160-168
Pangasianodon hypopthalmus	**E**	156-168 258-263	416-426	22-27 77-89 161 154
	O	**200-209** 256-268	471-495	38-40 81-86 156-166 170-177
Ictalurus furcatus	**E**	193-198 260-262	43 126-128 284-289	24-26 36 78-83 161 154
	O	201-207 264-271	43-45 132-135 286-293	* 37-39 84-87 157-163 166-172
Ictalurus punctatus	**E**	204 262-263	128-129 338	26-27 36 89 315
	O	199-208 263-274	131-137 327-342	* 36-39 84-89 316-333
Hybrid (*I. puncatus*)	**E**	197-199 261-267	127-130 331-338	25-28 36 82-84 314-315
	O	204-208 263-273	135-137 334-341	37-39 87-88 324-331

Table 7: Summary of identification matches based on each catfish species consensus sequence of cytochrome b region BLASTED against entries in GenBank database.

Species studied	GenBank (BLASTN)	
	Species identification	**% Max identity**
Ictalurus furcatus	*Ictalurus furcatus*	99
Ictalurus punctatus	*Ictalurus punctatus*	100
Hybrid (*I.punctatus* x *I. furcatus)*	*Ictalurus punctatus*	98
Clarias batrachus	*Clarias batrachus*	96
Clarias gariepinus	*Clarias gariepinus*	97
Clarias macrocephalus	*Clarias macrocephalus*	97
Pangasius bocourti	*Pangasius bocourti*	98
Pangasius hypophthalmus	*Pangasianodon hypophthalmus*	99

Figure 3: Computer generated Restriction Fragment Length Polymorphism (RFLP) gel image on 2100 Bioanalyzer for cytochrome b fragments of catfish species. Each gel image consists of the following enzyme profiles: a) *DdeI*, b) *Hae*III and c) *Nla*III. Two lanes of size markers (15 -1500 bp) were incorporated into each profiles.

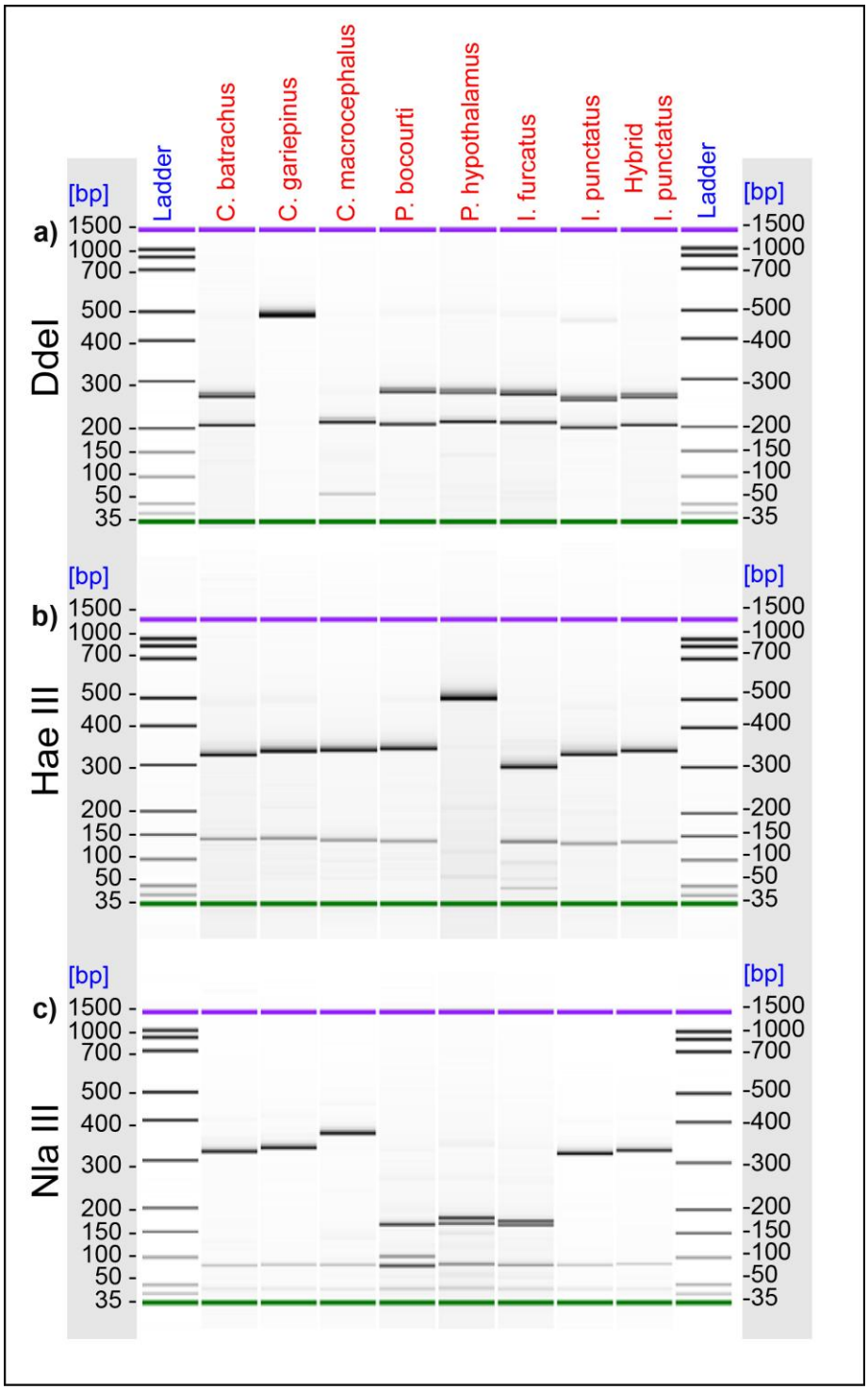


Figure 4: Electropherograms show the presence of species defining peaks (band size) for *Ictalurus punctatus* which corresponds to the RFLP profiles generated by three restriction endonucleases a) *Dde*I, b) *Hae*III and c) *Nla*III.

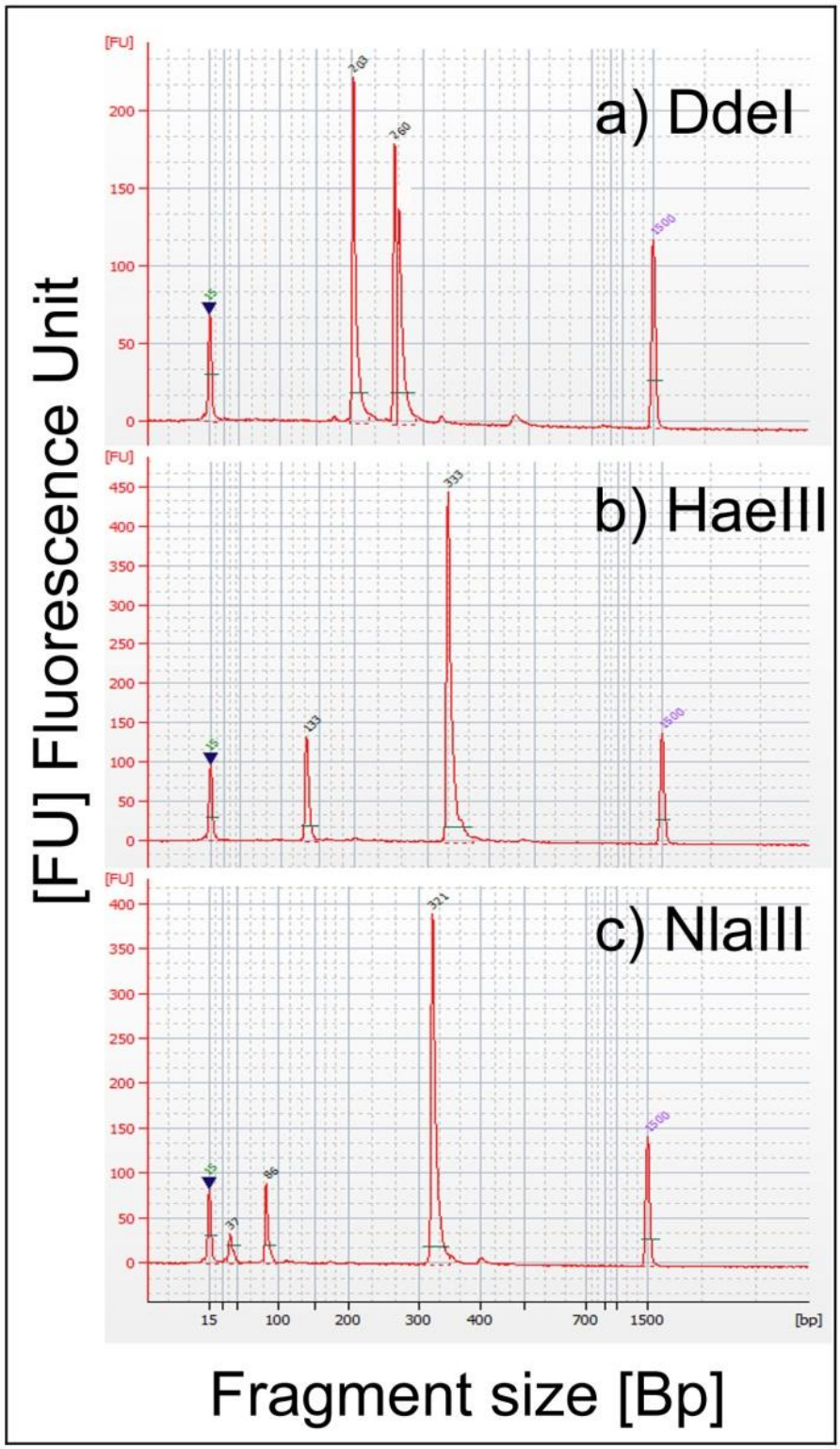


Figure 5: Species identification flowchart developed for seven catfish species (and an Ictalurid hybrid) based on the PCR-RFLP profiles generated by 2100 Bioanalyzer

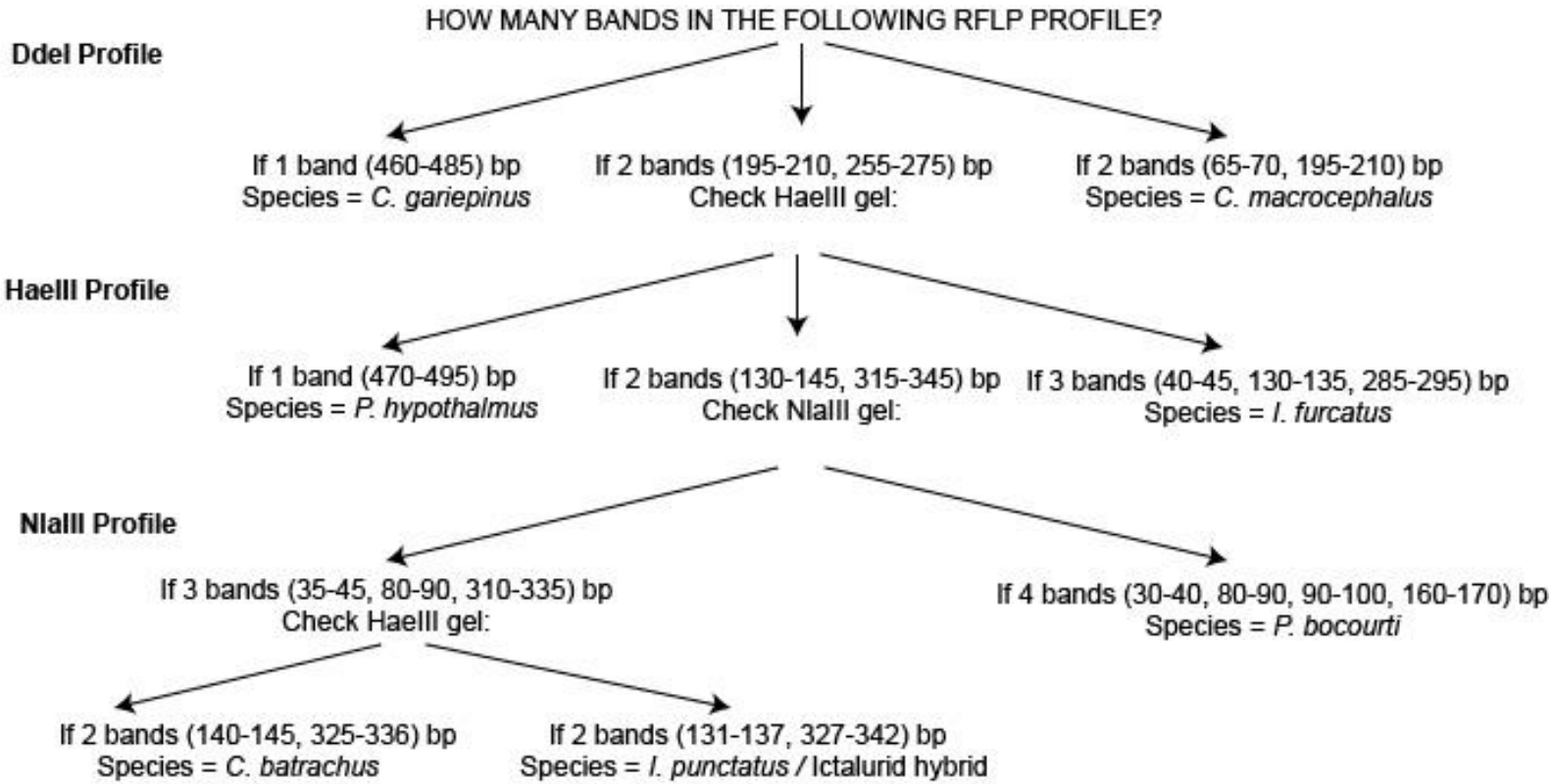
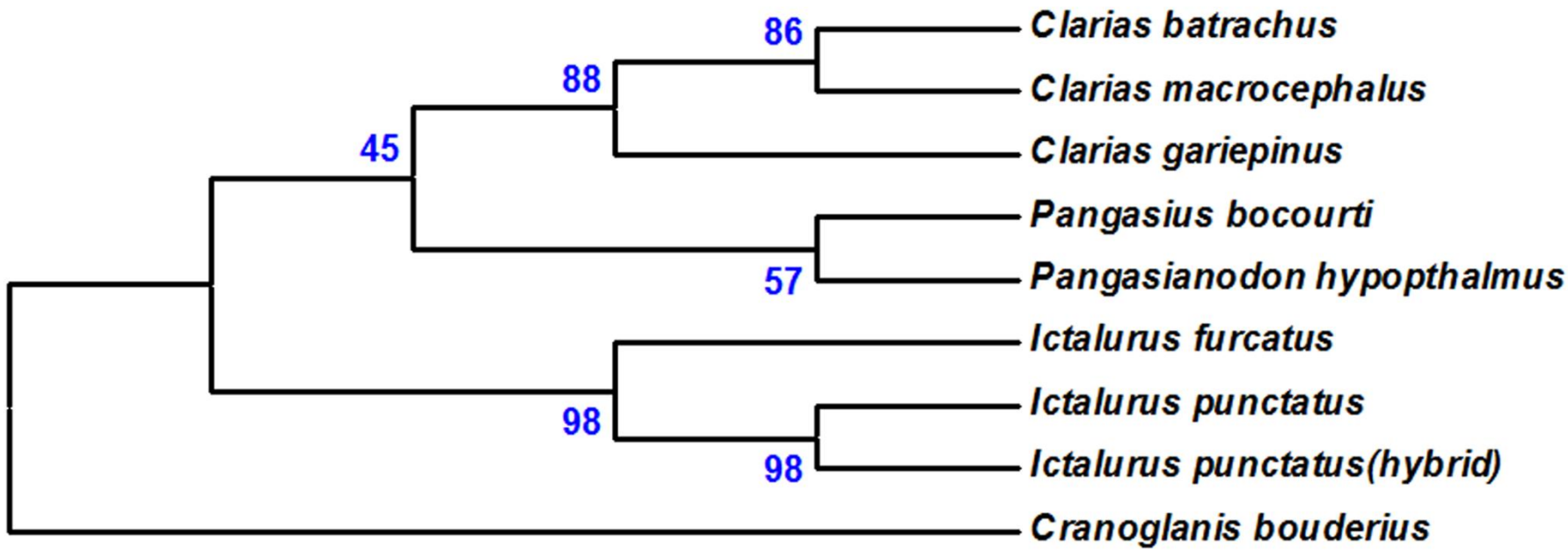


Figure 6: Phylogenetic consensus tree of seven catfish species (and an Ictalurid hybrid) constructed using Maximum Parsimony (MP) Method. Bootstrap values are shown next to the branches. The MP tree was built using the Close-Neighbor-Interchange algorithm.



CHAPTER 3

POPULATION STRUCTURE OF REDEYE BASS (*MICROPTERUS COOSAE*) IN ALABAMA

Abstract

The population structure of 451 redeye bass (*Micropterus coosae*) from 21 sampling sites in three main water drainages in Alabama was analyzed using 10 polymorphic microsatellite loci. Level of genetic diversity averaged across loci was consistently low in all populations with N_a ranging from 1.4 to 3.2. Substantial genetic differentiation was identified among the three major river basins with the highest divergence observed between Coosa and Cahaba Basin (Fst=0.256), suggesting restricted gene flow at the micro-geographical scale (river basin boundaries). Likewise, analysis of molecular variance (AMOVA) showed that most of the differentiation occurred between river basins, which constituted almost 50% of the total variance. Both STRUCTURE and phylogenetic analyses revealed four genetic clusters with substantial admixture events between water drainages. Overall, the patterns of genetic structuring were strongly associated with river basins rather than geographic distances, signifying the occurrence of human-mediated translocations. These genetic data provide informative input in the development of local management strategies of redeye bass populations in Alabama.

Background

Incidence of invasive species being reported as indigenous as well as native species being considered as exotic species are not uncommon (van Leeuwen et al. 2008; Castilla et al. 2002). Indeed, introduction of fish species either interspecifically or intraspecifically outside their native range has been documented to adversely threat the biodiversity of the indigenous community (Mack et al. 2000; Petit 2004). This may prove to be insidious, especially when the translocated populations are competitively superior, which could reduce the genetic diversity and subsequently the population size of the locally endemic populations (Kawamura et al. 2001). These effects could occur through food and spatial competition (Bøhn et al. 2008), predation (Valero et al. 2008) and transmission of diseases (Whittington and Chong 2007). In addition to these effects, human activity such as construction of artificial barriers (weirs and dams) also poses additional threats by impeding gene flow among the river systems, leading to contemporary fragmentation of the fish populations. Therefore, determination of the genetic structure between native and non-native populations is crucial before any introduction or eradication programs are implemented. In order to achieve this, an understanding of the genetic structures spanning the geographical locations at which the fish populations are distributed and differentiated as well as the determination of gene flow level can assist in the successful management of the conservation boundaries (Salgueiroa et al. 2003).

Redeye bass (*Micropterus coosae*) are black bass species that are endemic to the Southeastern United States and are typically found in cool streams and rivers in the foothills of mountains (Koppelman and Garrett 2002). Redeye bass have high aesthetic value (Hurst et al.

1975) and are readily distinguishable from other black basses by the appearance of white edges on the lower and upper part of their caudal fin.

Recent research has addressed the co-evolutionary relationship between redeye bass and freshwater mussels. Redeye bass has been identified to be an important reproductive host for two mussel species, *Hamiota altilis* and *Villosa vibex* (Haag et al. 1999). Both species are endemic to the river drainages where redeye bass are distributed, and *Hamiota altilis* is listed as endangered by the U.S. Fish and Wildlife Service (U.S. Fish and Wildlife Service 1994). Thus, proper resolution of phylogeographical status of redeye bass is crucial to the understanding of this co-evolutionary relationship, which is of utmost importance to the development of the mussels stocking plans in the future. Since previous genetic studies of redeye bass were primarily based on the hybridization studies, there is a lack of understanding of the species distribution throughout its native range. Although the population structures of several other black basses have been studied extensively, the population structure of redeye bass within its native range remains to be elucidated.

Different types of molecular marker were utilized to study population structure of black basses. Allozymes remained the primary genetic marker used to study population genetics of black basses from late 50's to even today. However, the availability of genomic tools and reagents, the discovery of mitochondrial sequences and microsatellites has changed the focus of molecular markers used to study population structures of these fish as well as other organisms. Due to its maternal mode of inheritance, mitochondrial DNA markers are often associated with the bias and erroneous estimation of species phylogeny, specifically when incidence of hybridization and introgression is inevitable (Flanders et al. 2009). Therefore, highly polymorphic nuclear markers such as microsatellites can readily compensate for this

disadvantage and became the most common marker used in population genetics studies (Heckel et al. 2005). In this study, a total of 451 redeye bass individuals sampled throughout their most prominently distributed locations in three main river basins in Alabama (Figure 7) were screened for genetic variation at 10 microsatellite loci. The present study represents the preliminary work performed to achieve the following aims: i) determination of the genetic structure of redeye bass populations among selected water drainages, ii) assessment of the level of population differentiation among the redeye bass populations in different tributaries of the selected water drainages. Upon achieving these goals, the long-term goal is aimed to generate information that eventually serves as essential guidelines for the resource management and conservation of redeye bass.

Materials and Methods

Sample collection

The sampling sites were chosen based on the historical distribution and abundance of redeye bass in water drainages in Alabama. Redeye bass were morphologically identified with the assistance of fisheries biologists from the Alabama Department of Conservation and Natural Resources. Number of sample collection varied according to availability of redeye bass in a given sampling site. The number of fish collected from all sampling sites ranged from 4 individual redeye bass to 45. Samples were collected by angling using baited hooks or by backpack electrofishing at the designated sampling sites (Figure 7). Total length (nearest mm) and weight (g) were measured and recorded for each fish. Geographical coordinations of each sample collection site were also recorded. After morphological identification, dorsal finclips

from each fish was taken non-lethally and preserved in 95% ethanol until further use. All redeye bass individuals were released back to their habitat after fin clip collection.

DNA extraction

Total genomic DNA was extracted from dorsal fin tissue of each fish following a digestion process using the protocol from Liu et al. 1998 and a modified version of the Puregene protocol for fish tissue (Qiagen, Valencia, CA). The integrity of the isolated DNA was visually examined on a 1.0% agarose gel with ethidium bromide while quantification of DNA concentration and purity was checked by standard spectrophotometry using Ultrospec 1100 Pro (GE Sciences, Pistachaway, NJ).

Microsatellite PCR optimization and genotyping

Ten primers (Table S3) representing microsatellite loci with different levels of polymorphism originally developed for other *Micropterus sp*. (Colbourne et al. 1996; Lutz-Carrillo et al. 2007; DeWoody et al. 2000; Malloy et al. 2000) were tested for suitability for redeye bass. Suitability was based on the amplification ability, production of consistent scorable banding patterns and level of polymorphism among individuals (Reid et al. 2008; Carrillo et al. 2006). Each microsatellite locus was either amplified individually or multiplexed by combinig 2 or 3 loci in a single PCR reaction. Information on amplification type, primer-specific annealing temperatures and PCR conditions were detailed in Table S3. All PCR reactions were performed in a total volume of 10µl mixture. The reaction mixture consisted of 1x PCR buffer (20 mm Tris-HCl pH 8.4, 50 mm KCl), 50-100 ng of genomic DNA, 1.5-2.0 mM $MgCl_2$, 0.025-0.5 µM of forward primer (contains a modified M13 tail at the 5' end), 0.08-0.50 µM of reverse primer,

0.20 μM of $IRD^{700/800}$ (infrared dye 700/800) labeled with modified M13 primer (sequence complimentary to the tail added to 5' end of the forward primer) (LI-COR, Lincoln, NE), 2.0 mM of each deoxynucleotide triphosphate (dNTP), 0.5 unit of Platinum Taq polymerase (Invitrogen, Carlsbad, CA) and distilled H_2O. The amplifications were carried out in a thermal cycler PTC-200 DNA Engine Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). Generally, the amplification profiles started with an initial denaturation step at 94°C for 2-5 min; repeated cycles of denaturation step at 94°C for 20s-30s, annealing step (T_a specified for each locus) for 10s-30s and elongation at 65°C / 72°C for 30s, and finalized with an extension step at 65°C / 72°C for 1-15 min. The amplified products were denatured at 95°C for 5 min and analyzed on an NEN 4300 Global IR2 DNA Sequencer (LI-COR, Lincoln, NE). A size standard of 50-350 bp was ran at the center and at both extremes of the gel to determine allele size as well as to ensure scoring accuracy and consistency across platforms.

Microsatellite Data Analysis

Prior to further population genetic and statistical analysis, Micro-Checker version 2.2.3 (van Oosterhout et al. 2004) was used to identify genotyping errors due to null alleles, stuttering or small allele dominance. Null alleles exist when mutation occur at primer sites causing non-amplification of certain alleles, which results in an excess of homozygotes (Pemberton et al. 1995). PCR stutter products are caused by slippage during PCR-amplification, which produce products that differ from the original amplicons by multiple of the repeat unit length (Shinde et al. 2003), impeding the discrimination between homozygotes and heterozygotes. Whereas small allele dominance or large allele dropout is inferred when small alleles are preferably amplified than larger alleles (Wattier et al. 1998). Analysis was performed with Bonferroni correction and

all scoring errors were adjusted based on Chakraborty's method (Chakraborty et al. 1992) so that the allele and genotype frequencies are readily used for further population genetic analysis.

A range of population genetics parameters was computed in GenAlEx version 6.2 (Peakall and Smouse, 2006), including the average number of alleles per locus (N_a), expected heterozygosity (H_e), observed heterozygosity (H_o) and fixation index (Fst). Pairwise Fst values and their corresponding significance level were verified by 999 permutations of individuals among samples (Schneider et al. 2000). Fst value ranges from 0-1 and higher Fst value indicates higher genetic differentiation among populations. An Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic tree was constructed based on pairwise Fst dataset using MEGA version 5.0 (Tamura et al. 2007).

Population genetic structure was investigated through a hierarchical analysis of molecular variance (AMOVA) with Arlequin version 3.11 (Excoffier et al. 2005). Microsatellite data were partitioned into four levels to enable a comparison of variation among basins, among populations within basins, among individual within populations and within individuals. For analysis between river basins, the variables were grouped as follows (Figure 7): Coosa Basin (sites A through J), Tallapoosa Basin (sites K through S), and Cahaba Basin (T and U). Under the assumption of stepwise mutation model, Rst (Slatkin 1995) was applied to estimate population differentiation through variation in allelic state (number of repeats) of microsatellite loci within and among populations.

Without a priori assumption of group membership, STRUCTURE version 2.3.3 (Pritchard et al. 2000), was used to detect the patterns of population differentiation by inferring the number of distinct genetic clusters as well as assigning individuals to these genetic clusters. Individuals were shuffled among populations and clustered them into the groups that

corresponded to maximum posterior probabilities of the data. Through this approach, the hierarchical relationships between populations can be restructured. This Bayesian based program was used to determine the potential number of genetic clusters in the dataset (K) and the affinity of individuals for these clusters, which is estimated as (Q). The likelihood of both of these values was estimated using the maximum likelihood approach, in which the true K value was determined following the criterion suggested by Evanno et al. (2005). Analysis was performed using the admixture model with correlated allele frequencies. The admixture model indicates that each individual has a proportion of genome from each of the K populations, whereas the correlated allele frequencies assume the presence of similarity between different populations due to migration or shared ancestry (Falush et al. 2003). To estimate the number of clusters, 2 independent simulations for each K (2-21) were run with a burn-in length of 100,000 iterations followed by 100,000 Markov Chain Monte Carlo (MCMC).

Results

Ten microsatellite loci were used for population structure analysis. Upon initial analysis using Micro-Checker with 451 individual samples indicated no evidence for large allele dropout for all loci tested. Only five loci (Mdo6, Msa13, Msa21, Mdo9 and Mdo3) showed the presence of null alleles as suggested by homozygote excess at these loci. Scoring errors due to stuttering also occurred at these loci. Despite that, all loci were in Hardy Weinberg equilibrium. A total of 37 alleles were generated from 10 microsatellite loci across 21 redeye bass populations (451 individuals). The size of microsatellite loci ranged from 120 to 222 bp, with low estimates of null allele frequencies for each locus across all populations. The average number of alleles per locus

(N_a) was consistently low in all populations ranging from 1.4 to 3.2 with lower values tending to occur in Tallapoosa Basin. The highest N_a was observed in Cheaha Creek (C) in Coosa basin (N_a=3.2) (Table 8). Observed heterozygosity (H_o) values ranged from 0.303 to 0.619 in Coosa Basin, 0.199 to 0.626 in Tallapoosa Basin and 0.167 to 0.278 in Cahaba Basin, with the highest H_o value in the Horseshoe Bend Creek (H_o=0.626). Among 10 microsatellite loci analyzed, Mdo3 had the largest number of alleles (n=9), followed by MiSa TPW134 (n=7) and four loci (Mdo2, Mdo9, MSA21, and LMA12) had only 4 alleles per loci (n=4).

The global estimate of Fst over all redeye bass populations and 10 microsatellite loci was significantly different from zero (Fst= 0.592, $P<0.05$) indicating a considerable population divergence. As shown in Table 9, the pairwise Fst values over all populations ranged from 0.014 (between Wadley and Upper Mainstem Tallapoosa) to 0.757 (between Big Willis Creek and Shoal Creek). Comparison of pairwise Fst values between basins revealed that the highest Fst value computed was between Coosa and Cahaba Basins (Fst=0.256), while the lowest Fst value was between Tallapoosa and Cahaba Basins (Fst=0.197), albeit a relatively low divergence value between them.

The phylogenetic tree reconstructed by UPGMA presented in Figure 8 revealed two different major branches. The first branch included 12 redeye bass populations (A-L), 10 of them from Coosa Basin. The other two populations from Crooked and Enitachopco Creek, collected from Tallapoosa Basin were clustered with populations from Coosa Basin. However, these two populations along with Walnut Creek sub-branched within the Coosa Basin branch. Allele frequencies from these two populations indicated that the frequencies at three loci, $Mdo1^{216}$, $Mdo7^{188}$, and $Mdo9^{140}$ were similar to corresponding allele frequencies observed in populations from Coosa Basin. The second major branch included redeye bass populations from

bothTallapoosa and Cahaba basins. Cahaba Basin populations also formed a separate sub-branch indicating allele frequency differences.

There were significant variation at all three hierarchical levels for the analysis of AMOVA, where populations were divided according to three main water drainages (Table 4). The highest level of differentiation was observed within individuals (48.53 %). Divergence between river basins was shown as 47.68% of the variation in the microsatellite data. The Garza-Williamson index estimated were relatively high which ranged from 0.36 to 0.59, signifying no signature of recent bottleneck events (Garza and Williamson 2001).

I have utilized the microsatellite data for admixture analysis using STRUCTURE version 2.3.3, a Bayesian based program that is able to partition individuals of unknown origin into different genetic groups. Following the simulation study by Evanno et al. (2005), the actual number of clusters, K is determined by plotting the K values of simulation analysis from 2 to 21 obtained from STRUCTURE against the values of logarithmic probability of data, L(K) (Pritchard et al. 2000). The curve of K versus the values of logarithmic probability of data plateaus once the estimated true K value is reached. In this study, the curve peaked at K=4 (L (K) = -7258.4), which strongly indicated the presence of four groups (Figure 9). These four groups were identified to represent the population structure of all studied sampling sites. However, increasing K to 5 did not substantially modify the population structuring pattern, suggesting that STRUCTURE was unable to segment the remaining genetic divergence. From this analysis, three distinct clusters partitioning Coosa Basin was observed. Despite the genetic grouping by STRUCTURE, populations in three basins appeared to have no correlation with geographical distance. In addition to that, populations (Figure 7) from 3 sample locations (K) of Tallapoosa Basin also clustered together with several populations from Coosa Basin (H-J) rather than the

geographically more proximate locations of the same water drainages (L-S). Interestingly, the fourth cluster was formed by geographically distant populations from both Tallapoosa Basin and Cahaba Basin (Figure 9). However, both populations from Cahaba Basin (T-U) showed evidence of admixture with predominant fourth cluster and two other clusters. Similarly, populations from Coosa Basin (D and G) and Tallapoosa Basin (L-M) were also markedly admixed with different clusters.

Discussion

This study utilized 10 polymorphic microsatellite loci to analyze the genetic structure of 21 redeye bass populations across the species geographic range in Alabama. My results consistently showed concordant patterns and highlighted substantial genetic distinctiveness between the river basins (% variation= 47.68), which suggested restricted gene flow at macro-geographic scale (among river basins) for redeye bass populations. Geographic distance can be used to explain this outcome, as species distribution is expected to be naturally discontinued since these populations were isolated and situated in different and distantly located regions.

Based on pairwise Fst values (Table 9), redeye bass populations within both Tallapoosa and Cahaba Basins appeared to be genetically similar. Although populations within a given basin are expected to be genetically similar to each other, redeye bass populations in Coosa basins however, displayed an unusual genetic structure. This was evident when geographically distant tributaries within Coosa Basin, Big Canoe Creek (A) and Big Willis Creek (B) were more genetically related than to their adjacent populations based on their pairwise Fst values, suggesting that gene flow has occurred between those populations.On the other hand, in the same

river basin, lower levels of genetic exchange was observed between Big Canoe Creek (A) and Little Canoe Creek (F) which have resulted in higher genetic differences between these populations albeit these two populations are in close proximity. Although this genetic pattern is relatively rare in natural populations, given the increased popularity of redeye bass as sport fish, it is predicted that human-mediated introduction into non-native habitat has created genetic admixture within these tributaries, which was also previously reported by Rohde et al. (1994).

Despite low genetic diversity (overall Ne =1.7, Table 9), most of the populations were highly genetically structured and have no signatures of bottleneck events owing to relatively high heterozygosity. However, similar to the outcome of Fst analysis, this genetic structuring was weakly related to geographic distance but strongly associated to the boundaries formed by river basins which act as stronger barriers to fish migration than spatial distance (Congdon 1995). Genetic variation and resolution of genetic markers for population structure studies are also highly influenced by demographic process. Therefore, reductions in population size would decrease the overall genetic variability of population. This implication can be responsible for the low number of alleles in following five populations (Tallapoosa Basin: O=Price Island Creek, P=Upper Mainstem Tallapoosa, Q=Wadley Creek; Cahaba Basin: T=Cahaba River, U=Little Cahaba).

There is compelling evidence for genetic structuring in *Micropterus coosae* populations based on our 10 microsatellite markers. Although AMOVA analysis has provided the basic information about the hierarchical genetic structure of redeye bass populations in Alabama, no indication of the clustering membership was given (Table 10). According to this analysis, high levels of differentiation were demonstrated between river basins due to the common feature of freshwater species, which has strong phylogeographic structure that is controlled by genetic drift

or selection (Thacker et al. 2007; Gum et al. 2005). On the other hand, STRUCTURE analysis found that there were relatively high levels of admixture among the river basins despite strong population differentiation. Some redeye populations were genetically diverged from populations within the same river basin but were genetically related to those from other basins. In fact, Crooked Creek (K) from Tallapoosa Basin was clustered with the remaining populations of Coosa Basin (H-J), rather than with populations from the more proximate locations in the Tallapoosa Basin (L-S). This weak isolation by distance event displayed by the admixture of populations between different river basins most likely reflects the anthropogenic dispersal of redeye bass across major river systems in Alabama. Similarly, admixture genetic structure was also observed in UPGMA phylogenetic tree, which further supported this inference.

Although the sample sizes of redeye bass populations in this study might be sufficient to determine the genetic structure of *Micropterus coosae* in major water drainages system in Alabama, additional research including more microsatellite loci and larger sample sizes covering a larger geographical scale should be conducted. This would provide more accurate estimates of gene flow between river systems as well as insight into determination of management units (MUs). Additionally, comparisons of morphometric, biological and ecological data (breeding patterns, habitat preference, sexual isolation, local adaptation etc.) will also facilitate the clarification of taxanomic status, which is crucial for conservation and management of *Micropterus coosae*, especially in admixture populations or tributaries with lower genetic diversity. Through deep understanding of the population structure, successful introduction and eradication programs could be implemented to preserve the biodiversity of redeye bass as well as other species that co-evolution with them. For instance, isolated redeye bass populations with high levels of genetic differentiation can be potential targets for eradication efforts. With these

information and expanded genetic study, appropriate conservation management boundaries and protection areas such as evolutionary significant units (ESUs) and management units (MUs) can be identified (Hey et al. 2003; Wan et al. 2004). Accordingly, populations with dissimilar genetic structure should be considered as separate ESUs, in which translocation of individuals between them should be prohibited (Moritz 1994). The results of my study suggest that redeye bass populations from three river basins in Alabama appeared to have low genetic diversity but are highly structured. My results from STRUCTURE and phylogenetic analysis could be used as a guideline to define the management unit of redeye bass populations. Overall, this study is anticipated to provide basis for the conservation and management of *Micropterus coosae* as well as their co-evolutionary mussel species in Alabama.

References

1. Bøhn T, Amundsen PA and Sparrow A (2008). Competitive exclusion after invasion? Biology Invasions 10:359-368.
2. Carrillo DJL, Nice CC, Bonner TH and Forstner MRJ (2006). Admixture Analysis of Florida Largemouth Bass and Northern Largemouth Bass using Microsatellite Loci. Transactions of the American Fisheries Society 135: 779-791.
3. Castilla JC, Collins AG, Meyer CP, Guinez R and Lindberg DR (2002). Recent introduction of the dominant tunicate *Pyura praeputialis* (Urochordata, Pyuridae) to Antofagasta, Chile. Molecular Ecology 11:1579-1584.
4. Chakraborty R, De Andrade M, Daiger SP and Budowle B (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Annals of Human Genetics 56:45-57.
5. Colbourne JK, Neff BD, Wright JM and Gross MR (1996). DNA fingerprinting of bluegill sunfish (*Lepomis macrochirus*) using (GT) n microsatellites and its potential for assessment of mating success. Canadian Journal of Fisheries and Aquatic Sciences 53:342-349.
6. Congdon BC (1995). Unidirectional gene flow and maintenance of genetic diversity in mosquitofish *Gambusia holbrooki* (Teleostei: Poeciliidae). Copeia: 162-172.
7. DeWoody JA, Fletcher DE, Wilkins SD, Nelson WS and Avise JC (2000). Genetic monogamy and biparental care in an externally fertilizing fish, the largemouth bass (*Micropterus salmoides*). Proceedings of Royal Society B 267: 2431-2437.
8. Excoffier L, Laval G and Schneider S (2005). Arlequin ver. 3.1: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.
9. Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
10. Flanders J, Jones G, Benda P et al. (2009). Phylogeography of the greater horseshoe bat, *Rhinolophus ferrumequinum*: contrasting results from mitochondrial and microsatellite data. Molecular Ecology 18: 306-318.
11. Falush D, Stephens M, and Pritchard JK (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics 164:1567-1587.
12. Garza JC and Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. Molecular and Ecology 10: 305-318.
13. Gum B, Gross R and Kuehn R (2005). Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. Molecular Ecology 14: 1707-1725.
14. Haag WR, Warren ML and Shillingsford M (1999). Host fish and host-attracting behavior of *Lampsilis altilis* and *Villosa vibex* (Bivalvia: Unionidae).The American Midland Naturalist 141: 149-157.
15. Heckel G, Burri R, Fink S, Desmet J and Excoffier L (2005). Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution 59: 2231-2242.
16. Hey J, Waples RS, Arnold ML, Butlin RK and Harrison RG (2003). Understanding and confronting species uncertainty in biology and conservation. Trends in Ecology and Evolution 18:597-603.

17. Hurst H, Bass G and Hubbs C (1975). The biology of the Guadulape, Suwannee, and redeye basses. In H Clepper (ed). Black bass biology and management. Sport Fishing Institute, Washington D.C. 47-53 pp.
18. Kawamura K, Ueda T, Arai R et al. (2001). Genetic introgression by the rose bitterling, *Rhodeus ocellatus ocellatus*, into the Japanese Rose Bitterling, *R-o kurumeus* (Teleostei, Cyprinidae). Zoological Science 18: 1027-1039.
19. Koppelman J and Garrett G (2002). Distribution, biology and conservation of the rare black bass species. In D Phillip and M Ridgway (eds). Black bass: ecology, conservation and management. American Fisheries Society 31: Bethesda, Maryland.
20. Liu Z, Nichols A, Li P and Dunham RA (1998). Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*Ictalurus furcatus*) and their F_1, F_2 and backcross hybrids. Molecular and General Genetics 258: 260-268.
21. Lutz-carrillo D, Hagen C, Dueck LA and Glenn TC (2007). Isolation and characterization of microsatellite loci for Florida largemouth bass, *Micropterus salmoides floridanus*, and other micropterids. Molecular Ecology Notes. doi: 10.1111/j.1471-8286.2007.01917.x
22. Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M and Bazzaz FA (2000). Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Application 10:689-710.
23. Malloy TP, Vandenbussche RA, Coughlin WD and Echelle AA (2000). *Micropterus dolomieu* (Teleostei: Centrarchidae), and cross-species amplification in spotted bass, *M. punctulatus*. Molecular Ecology 9:1919-1952.
24. Moritz C (1994). Defining 'Evolutionary Significant Units' for conservation. Trends in Ecology and Evolution 9: 373-375.
25. Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.
26. Pemberton JM, Slate J, Bancroft DR and Barrett JA (1995). Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Molecular Ecology 4: 249-252.
27. Petit RJ (2004). Biological invasions at the gene level. Diversity and Distributions 10: 159-165.
28. Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.
29. Reid SM, Wilson CC, Mandrak NE and Carl LM (2008). Population structure and genetic diversity of black redhorse (*Moxostoma duquesnei*) in a highly fragmented watershed. Conservation Genetics 9: 531-546.
30. Rohde F, Arndt R, Lindquist D and Parnell J (1994). Freshwater Fishes of the Carolinas, Virginia, Maryland, and Delaware. The University of North Carolina Press, Chapel Hill, North Carolina.
31. Salgueiroa P, Carvalho G, Collares-Pereiraa MJ and Coelhoa MM (2003). Microsatellite analysis of genetic population structure of the endangered cyprinid *Anaecypris hispanica* in Portugal: Implications for Conservation. Biological Conservation 109:47-56.
32. Schneider S, Roessli D and Excoffier L (2000). ARLEQUIN: Software for population genetics data analysis v 2.0. Department of Anthropology, University of Geneva, Switzerland.

33. Shinde D, Lai YL, Sun FZ and Arnheim N (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: $(CA/GT)_n$ and $(A/T)_n$ microsatellites. Nucleic Acid Research 31: 974-980.
34. Slatkin M and Excoffier L (1996). Testing for linkage disequilibrium in genotypic data using the EM algorithm. Heredity 76: 377-383.
35. Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.
36. Thacker CE, Unmack PJ, Matsui L and Rifenbark N (2007). Comparative phylogeography of five sympatric *Hypseleotris* sp (Teleostei: Eleotridae) in south-eastern Australia reveals a complex pattern of drainage basin exchanges with little congruence across species. Journal of Biogeography 34:1518-1533.
37. US Fish and Wildlife Service (1994). Endangered and threatened wildlife and plants. Division of Endangered Species, US Fish and Wildlife Service, Washington DC. 42 pp.
38. Valero A, Macı´as Garcia C and Magurran AE (2008). Heterospecific harassment of native endangered fishes by invasive guppies in Mexico. Biology Letter 4:149-152.
39. Van Leeuwen JFN, Froyd CA, van der Knaap WO, Coffey EE, Tye A and Willis KJ (2008). Fossil pollen as a guide to conservation in the Galapagos. Science 332: 1206.
40. Van Oosterhout C, Hutchinson WF, Wills DPM and Shipley P (2004). MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538. doi: 10.1111/j.1471-8286.2004.00684.x
41. Wan QH, Wu H, Fujihara T and Fang SG (2004). Which genetic marker for which conservation genetics issue? Electrophoresis 25:2165-2176.
42. Wattier R, Engel CR, Saumitou-Laprade P and Valero M (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology 7: 1569-1573.
43. Whittington RJ and Chong R (2007). Global trade in ornamental fish from an Australian perspective: the case for revised import risk analysis and management strategies. Preventive Veterinary Medicine 81:92-116.

Table 8: Sample sizes and summary of genetic diversity parameters of the redeye bass at 10 microsatellite loci sampled in Alabama. N, number of individuals genotyped at each locus; N_a, average number of alleles per locus; N_e, the number of effective alleles; H_o, observed heterozygosity; H_e, expected heterozygosity. Bolded values represent the lowest and highest values in each category (see text).

Sampling sites	N	N_a	N_e	H_o	H_e
Big Canoe Creek	24	2.8	1.817	0.483	0.343
Big Willis Creek	21	2.0	1.691	0.495	0.279
Cheaha Creek	32	**3.2**	2.317	**0.619**	0.465
Choccolocco Creek	21	2.6	2.158	0.586	0.443
Hatchet Creek	33	2.6	1.851	0.427	0.345
Little Canoe Creek	38	2.1	1.705	0.542	0.313
Little River	45	2.7	2.009	0.586	0.433
Little Willis Creek	17	1.7	1.585	0.529	0.296
Terrapin Creek	37	2.3	1.790	0.554	0.366
Walnut Creek	31	2.2	1.602	**0.303**	0.241
Mean		2.4	1.852	0.513	0.352
Crooked Creek	15	1.6	1.459	0.273	0.188
Enitachopco Creek	23	1.9	1.587	0.470	0.290
Horseshoe Bend Creek	23	2.5	1.854	**0.626**	0.408
Mad Indian Creek	10	1.7	1.646	0.550	0.306
Price Island Creek	6	**1.4**	1.400	0.400	0.200
Upper Mainstem Tallapoosa	5	**1.4**	1.347	0.340	0.182
Wadley Creek	4	**1.4**	1.346	0.300	0.159
Cohobadiah Creek	28	2.1	1.580	0.221	0.228
Shoal Creek	23	1.8	1.384	**0.199**	0.190
Mean		1.8	1.512	0.375	0.239
Cahaba River	9	1.6	1.446	**0.167**	0.194
Little Cahaba	6	2.7	1.898	**0.278**	0.391
Mean		2.2	1.672	0.222	0.293

Table 9: Pairwise Fst estimates (below diagonal) and their P-values (above diagonal) among all locations as detailed in Figure 7. Alphabets correspond to each redeye populations (see Figure 7). Color indicates the three different basins, red, Coosa; grey, Tallapoosa, and blue, Cahaba. Bold numbers indicate both the lowest and highest Fst values as described in the text.

*	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
A	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
B	0.177	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
C	0.249	0.211	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
D	0.364	0.361	0.173	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
E	0.484	0.491	0.364	0.215	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
F	0.532	0.555	0.399	0.330	0.167	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
G	0.488	0.510	0.376	0.295	0.190	0.175	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
H	0.597	0.613	0.500	0.458	0.384	0.455	0.250	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
I	0.577	0.588	0.485	0.426	0.350	0.450	0.297	0.157	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
J	0.686	0.712	0.589	0.558	0.553	0.620	0.428	0.389	0.243	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
K	0.698	0.745	0.597	0.593	0.600	0.625	0.451	0.492	0.378	0.276	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
L	0.658	0.699	0.570	0.563	0.568	0.563	0.432	0.481	0.380	0.350	0.212	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
M	0.609	0.644	0.511	0.521	0.541	0.558	0.402	0.450	0.344	0.369	0.372	0.243	***	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
N	0.652	0.704	0.532	0.565	0.615	0.619	0.487	0.587	0.483	0.532	0.576	0.428	0.161	***	0.050	0.080	0.080	0.010	0.010	0.010	0.010
O	0.680	0.738	0.559	0.598	0.648	0.641	0.491	0.622	0.545	0.613	0.657	0.517	0.247	0.128	***	0.240	0.180	0.010	0.010	0.010	0.010
P	0.681	0.741	0.559	0.597	0.650	0.645	0.497	0.634	0.557	0.631	0.683	0.539	0.255	0.149	0.039	***	0.190	0.010	0.010	0.010	0.010
Q	0.682	0.744	0.560	0.597	0.652	0.648	0.503	0.644	0.566	0.644	0.703	0.558	0.275	0.192	0.122	**0.014**	***	0.020	0.010	0.010	0.010
R	0.703	0.743	0.612	0.647	0.680	0.668	0.556	0.668	0.606	0.669	0.686	0.600	0.354	0.291	0.202	0.132	0.108	***	0.010	0.010	0.010
S	0.719	**0.757**	0.622	0.652	0.706	0.707	0.589	0.708	0.649	0.719	0.746	0.664	0.461	0.480	0.466	0.443	0.453	0.264	***	0.010	0.010
T	0.671	0.719	0.558	0.583	0.663	0.673	0.536	0.667	0.609	0.723	0.753	0.661	0.490	0.591	0.643	0.651	0.669	0.574	0.457	***	0.010
U	0.583	0.613	0.489	0.489	0.544	0.571	0.459	0.550	0.497	0.619	0.624	0.554	0.427	0.519	0.535	0.532	0.526	0.547	0.556	0.368	***

Table 10: Analysis of molecular variance (AMOVA) based on Rst distance method comparing the genetic variation in 10 microsatellite loci for *Micropterus coosae* from three main water drainages. This model compares genetic variation among river basins (Coosa, Tallapoosa and Cahaba), among populations within river basins (as in Figure 7, Coosa: A-J; Tallapoosa: K-S; Cahaba: T-U), among individuals within populations, and among 451 individuals.

Source of variation	*df*	Sum of squares	Variance components	Fixation indices	*P*-value	Percentage of variation
Among river basins	2	35239.3	773.9	0.477	<0.05	47.68
Among populations within river basins	18	28993.8	36.7	0.429	<0.05	22.46
Among individuals within populations	430	7849.7	-30.5	-0.626	1.000	-18.68
Within individuals	451	35754.0	79.3	0.515	<0.05	48.53
Total	901	107836.8	163.3	--	--	100.00

Figure 7: Schematic representation of sampling sites and letter codes of redeye bass populations collected from three river basins. River basins are indicated with different color shading. Actual sampling sites were indicated with a closed circle.

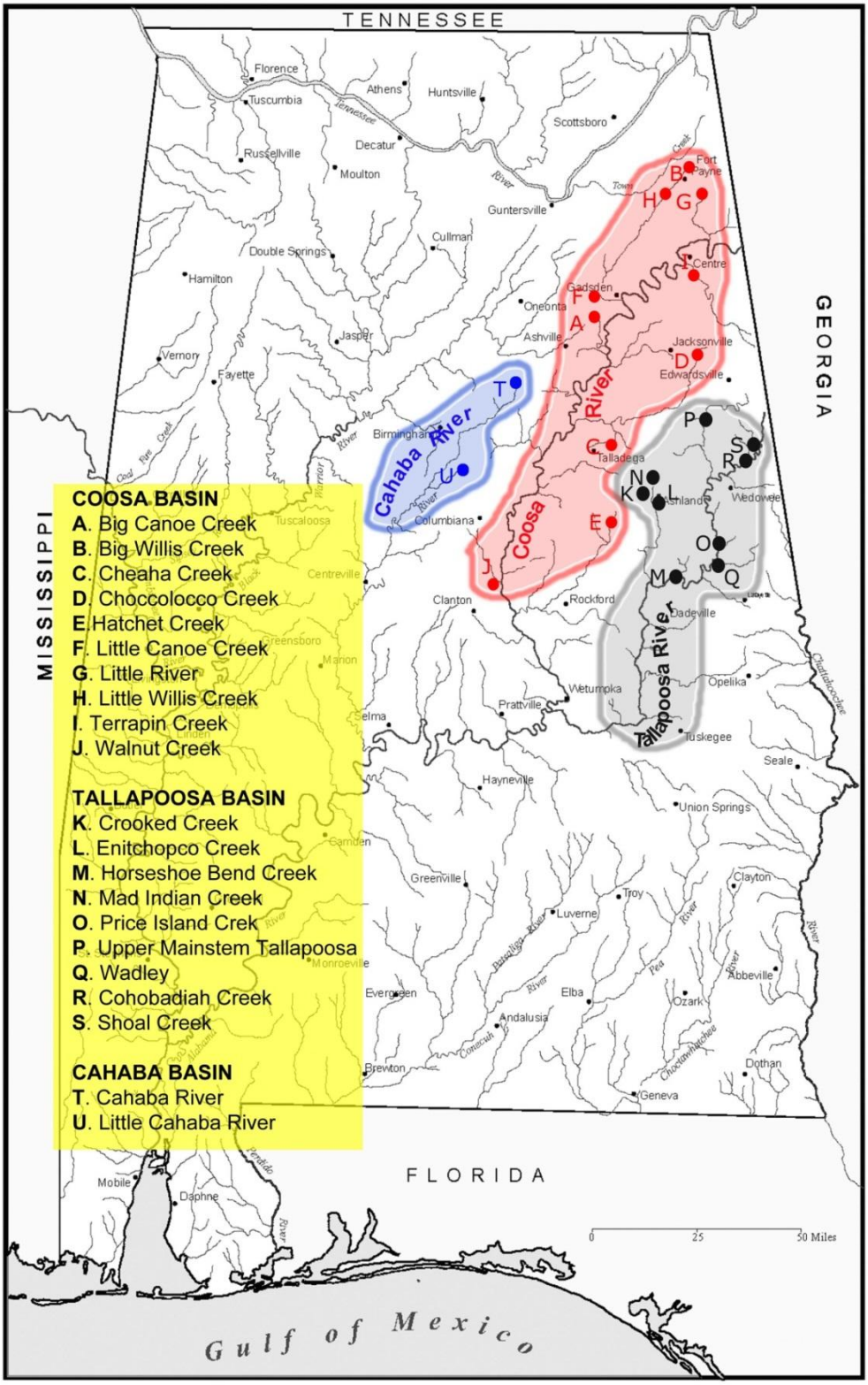


Figure 8: Genetic relationships of redeye bass populations based on the pairwise genetic differentiation index Fst values calculated from the microsatellite genotypes. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic tree reconstructed using MEGA 5.0 (Tamura et al. 2007). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

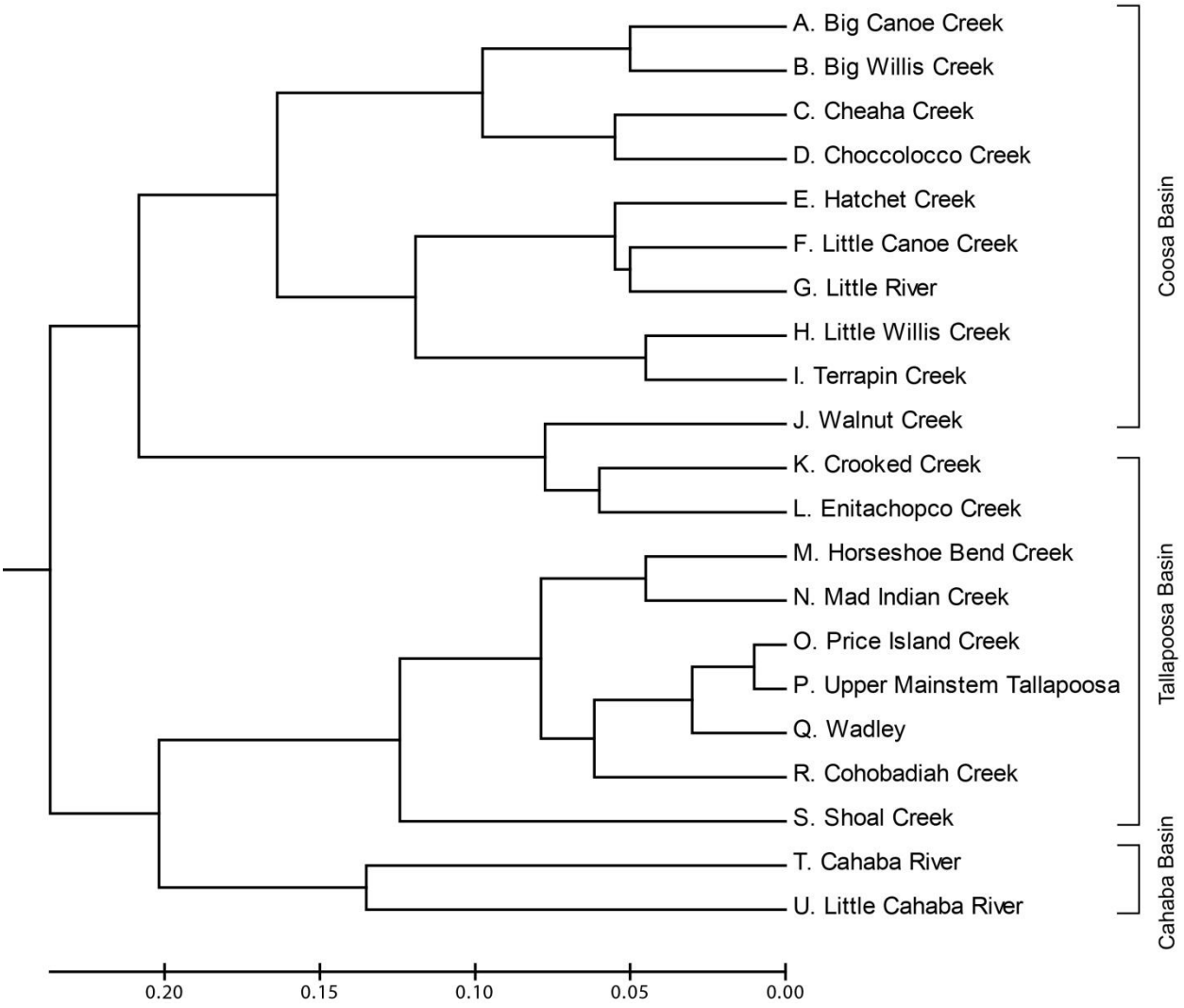


Figure 9: Summary plots of the estimated membership coefficient for each redeye bass individual from all populations based on 10 microsatellite loci. Four different color segments and solid lines under the chart represent 4 genetic clusters (K=4). Each individual is represented by single vertical bars partitioned into different colored segments which are proportional to the membership coefficient of genetic clusters. Each population is represented by a single capital letter (see Figure 7). STRUCTURE analysis indicated that population K (Crooked Creek) should have been pooled with the other populations in Tallapoosa River (L through S) instead of Coosa River (H through J).

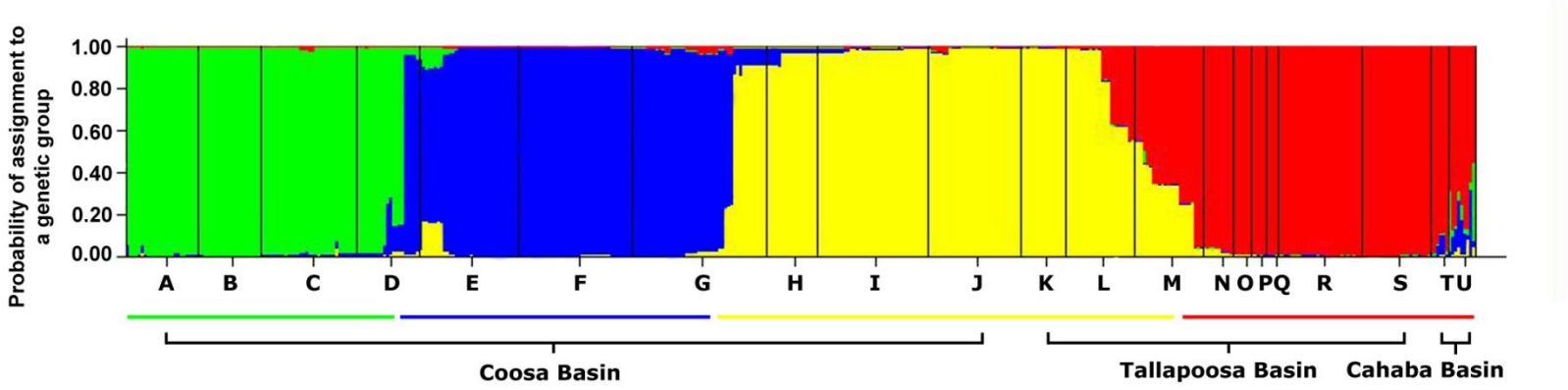


Table S3. Microsatellite primer sequences, PCR conditions, amplicon sizes and sources used in this study for *Micropterus coosae* populations in Alabama. The locus name listed included both forward and reverse primer sequences.

Locus	Primer sequence (5'-3')	Ta (°C); cycles number	Size range	Source
Mdo9	F: *GACCGGTCCTGCATATGATT R: TTTGATGGGCGTTTTGTGTA	55; 35	138-150	Malloy et al. 2000
LMA12	F: *TTCTTCCACAATATTCTCGCG R: CTGCTCAGCATGGAGGCAG	43; 35	120-132	Colbourne et al. 1996
MiSaTPW134	F: *AGATAACTCTTAGACACCAGTCCC R: TATGGTGGCGTGTTAGTGTGACG	57; 15 / 53; 10	154-166	Lutz-Carrillo et al. 2007
Mdo7	F: *GTCACTCCCATCATGCTCCT R: TCAAACGCACCTTCACTGAC	47.6; 35	176-196	Malloy et al. 2000
MSA21	F: *GTTGTCAAGTCGTAGTCCGC R:CACTGTAAATGGCACCTGTGG	47.6; 35	206-214	DeWoody et al. 2000
Mdo6	F: *TGTGTGGGTGTTTATGTGGG R: TGAAATGTACGCCAGAGCAG	45; 35	156-166	Malloy et al. 2000
MSA13	F: *CTTCTGTCCTGCATCCTCTTAG R: CTGATACAGCAGCTCGAAGC	45; 35	198-210	DeWoody et al. 2000
Mdo1	F: *ATCTCAGCCCATACCGTCAC R: GCTCTTCCCAGTGGTGAGTC	60; 6 / 58; 9 / 56; 3 / 45; 17	212-222	Malloy et al. 2000
Mdo2	F: *CTGCTCTGGCGTACATTTCA R: GCCCTTTCATATTGGGACAA	60; 6 / 58; 9 / 56; 3 / 45; 17	206-212	Malloy et al. 2000
Mdo3	F: *CTGCATGGCTGTTATGTTGG R: AGGTGCTTTGCGCTACAAGT	60; 6 / 58; 9 / 56; 3 / 45; 17	122-146	Malloy et al. 2000

* Modified M13 5'-GAGTTTTCCCAGTCACGAC-3'added to the 5' end of each forward primer sequence.

GENERAL DISCUSSIONS AND CONCLUSIONS

The advancement of next generation sequencing and other fingerprinting technologies has revolutionized how genetic analyses are carried out in various biological studies ranging from species identification (Botti and Giuffra 2010; Lowenstein et al. 2010), forensic sciences (Dawnay et al. 2007), molecular systematics and phylogenetics (Hardman 2005) to population genetic structure (Carillo et al. 2006; Flanders et al. 2009). My dissertation includes different types of analyses encompassing both analysis of an important aquaculture species from the food safety perspective and an important sport fish at the population levels from various biological and ecological issues. The common link of these independent projects was to use molecular techniques for the differentiation of fish at population or species levels.

In my dissertation, two separate studies using different DNA-based markers were performed to authenticate commercialized catfish species due to growing concern in terms of food safety involving increased seafood fraud. The first study focused on the development and validation of DNA barcoding using cytochrome oxidase I (COI) as an invaluable regulatory tool for catfish species identification. Consensus barcodes of 9 species (and an Ictalurid hybrid) represented by catfish families of Ictaluriidae, Clariidae, Pangasiidae, Cranoglanididae and Bagridae were generated. These sequences were then compared to the records in major databases (GenBank and BOLD) to determine whether congeneric variation between these species were sufficient for species differentiation. Even though one of the species studied (*Clarias batrachus*) failed to be perfectly matched with the recorded sequences of the databases, informative causes for this discrepancy were found, revealing the erroneous features and shortcomings in these two major databases. The lack of transparency and access to sequence tracefiles deposited in these

databases hinder the efforts to ascertain the accuracy of the identified samples. Nevertheless, my results including a blinded sample test indicated that DNA barcoding is a powerful technique, accurately identifying samples regardless of sample source.

The second study involved the development of Restriction Fragment Length Polymorphism (RFLP) profiles using microchip-based capillary electrophoresis system for the discrimination of commercialized catfish species. In spite of the fact that PCR and RFLP techniques have been well documented for the differentiation of species, the coupling of the two techniques, along with the adoption of a lab-on-chip platform, rather than the traditional gel electrophoresis, and their application to species differentiation for addressing catfish food safety brought practical importance to this work. My results indicated that RFLP profiles that were derived from a combinational of all three enzymes were able to produce restriction fragments patterns potentially to differentiate catfish species. In addition, other restriction enzymes can be tested in the future to assure full power of differentiation, and this work laid the ground work for continuation of this line of research. The application of lab-on-chip technology is attractive because compared to conventional gel-based electrophoresis, this endpoint detection method using 2100 Bioanalyzer allows higher speed and sensitivity of detection, improved resolution of RFLP patterns, usage reduction of harmful solvents and products (Doodey et al. 2005) as well as ease of handling. With this manifold of advantages, I believe that this method will be useful for the authentication of catfish species for the regulatory agencies.

The population level study involved analysis of the genetic variation patterns of redeye bass (*Micropterus coosae*) populations in the main water drainages in Alabama. Redeye bass is an important and highly desired sport fish in the Southeastern United States. In this study, using 10 microsatellite loci the genetic structure of redeye bass individuals from 21 sampling sites was

studied (Figure 7). My results showed concordant genetic patterns and highlighted substantial genetic differentiation of redeye bass populations among the three major river basins with the highest divergence observed between Coosa River Basin and Cahaba River Basin (Fst=0.256), indicating restricted gene flow at the micro-geographical scale (river basin boundaries). Interestingly, the pattern of genetic structuring was strongly associated with river basins rather than geographic distance. Geographically distant redeye populations within a river basin appeared to have similar genetic structure, whereas redeye populations within a close proximity within the same river basin displayed genetic divergence. Although this genetic pattern is relatively rare in natural populations, given the increased popularity of redeye bass as a sport fish, I hypothesized that human-mediated introduction into non-native habitat has created genetic admixture within these tributaries, as previously concluded with other species (Rohde et al. 1994). Phylogenetic tree reconstructed from the pair-wise Fst values clearly represented the genetic structuring of redeye bass populations in Alabama. This was further verified when STRUCTURE analysis revealed four genetic clusters with substantial admixture events between water drainages. Both STRUCTURE and phylogenetic tree analysis could be used as a guideline to define local management units for the conservation and management of *Micropterus coosae* as well as their co-evolutionary freshwater mussel species in Alabama.

As a whole, all three types of studies conducted using molecular markers, the COI seqences for barcoding, PCR-RFLP and microsatellites, were able to reveal variations within DNA sequences in fish. Apparently, these molecular techniques will be useful for genetic differentiation at various levels, perhaps more powerful with microsatellites to be able to differentiate at the population and even individual levels, whereas mt DNA is more useful for species level or population level studies.

REFERENCES

1. Abdo Z and Golding GB (2007). A step toward barcoding life: a model based, decision theoretic method to assign genes to preexisting species groups. Systems Biology 56:44-56.
2. Akasaki T, Yanagimoto T, Yamakami K, Tomonaga H, and Sato S (2006). Species Identification and PCR-RFLP Analysis of Cytochrome b Gene in Cod Fish (Order Gadiformes) Products. Journal of Food Science 71 (3): 190-195.
3. Allendorf F and Lundquist L (2003). Introduction: population biology, evolution and control of invasive species. Conservation Biology 17:24-30.
4. Arahishi F (2005). PCR-RFLP analysis of nuclear nontranscribed spacer for mackerel species identification. Journal of Agricultural and Food Chemistry 53:508-511.
5. Arif IA and Khan HA (2009). Molecular markers for biodiversity analysis of wildlife animals: a brief review. Animal Biodiversity and Conservation 32(1): 9-17.
6. Asensio L, Gonzalez I, Fernandez A, Rodriguez MA, Lobo E, et al. (2002). Application of random amplified polymorphic DNA (RAPD) analysis for identification of grouper (*Epinephelus guaza*), wreck fish (*Polyprion americanus*), and Nile perch (*Lates niloticus*) fillets. Journal of Food Protection 65: 432-435.
7. Avise J, Pearce P, Avyle MVD, Smith M, Nelson W and Asmussen M (1997). Cytonuclear introgressive swamping and species turnover of bass following an introduction. Journal of Heredity 84: 14-20.
8. Banhos A, Hrbek T, Gravena W, Sanaiotti T, and Farias IP (2008). Genomic resources for the conservation and management of the harpy eagle (*Harpia harpyja*, Falconiformes, Accipitridae). Genetics and Molecular Biology 31: 146-154.
9. Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, et al. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of ‘‘palombo”. *Mustelus spp*. Food Research International 43: 376-381.
10. Bartlett SE and Davidson WS (1991). Identification of Thunnus tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes. Canadian Journal of Fisheries and Aquatic Sciences 48: 309-317.
11. Barwick D and Moore P (1983). Abundance and growth of redeye bass in two South Carolina reservoirs. Transactions of the American Fisheries Society 112: 216-219.
12. Barwick H, Oswald K, Quattro J, and Barwick R (2006). Redeye bass (*Micropterus coosae*) and Alabama spotted bass (*M. punctulatus henshalli*) hybridization in Keowee reservoir. Southeastern Naturalist 5:66 -668.
13. Brown LR and Moyle PB (1997). Invading species in the Eel River, California: Successes, failures, and relationships with resident species. Environmental Biology Fishery 49: 271-291.
14. Brown LR and Ford T (2002). Effects of flow on the fish communities of a regulated California river: Implications for managing native fishes. River Restoration Application18: 331-342.
15. Boschung H and Mayden R (2004). Fishes of Alabama. Smithsonian Press, Washington, D.C.
16. Bossier P (1999). Authentication of seafood products by DNA patterns. Journal of Food Science 64: 189-193.
17. Botti S and Giuffra E (2010). Oligonucleotide indexing of DNA barcodes: Identification of tuna and other Scombrid species in food products. BMC Biotechnology 10: 60.

18. Brambilla I, Porto G and Tarozzi A (2009). Adjusting to Trade-Policy Changes in Export Markets Evidence from U.S. Antidumping Duties on Vietnamese Catfish. The World Bank Development Research Group Trade and Integration Team. Policy Research Working Paper WPS4990.
19. BurgessWE (1989). An Atlas of Freshwater and Marine Catfishes. A Preliminary Survey of the Siluriformes. Neptune City, New Jersey: Berkshire TFH Publications.
20. Butler GL (2009). Biology and conservation of the endangered eastern freshwater cod, *Maccullochella ikei*. PhD Thesis, Southern Cross University, Lismore, Australia.
21. Carrillo DJL, Nice CC, Bonner TH and Forstner MRJ (2006). Admixture Analysis of Florida Largemouth Bass and Northern Largemouth Bass using Microsatellite Loci. Transactions of the American Fisheries Society 135: 779-791.
22. Cathey HJ (1973). A study of the redeye bass *Micropterus coosae*, smallmouth bass *Micropterus dolomieu* and rockbass, *Ambloplites rupestris* in Roaring River, Tennessee. Unpublished master's thesis. Tennessee Technological University, Cookeville, Tennessee. 45 pp.
23. Cespedes A, Garcia T, Carrera E, Gonzalez I, Sanz B, et al. (1998). Polymerase chain reaction-restriction fragment length polymorphism analysis of a short fragment of the cytochrome b gene for identification of flatfish species. Journal of Food Protection 61: 1684-1685.
24. Cespedes A, Garcia T, Carrera E, Gonzalez I, Fernandez A, et al. (1999). Application of polymerase chain reaction-single strand conformational polymorphism (PCR-SSCP) to identification of flatfish species. The Journal of AOAC International 82: 903-907.
25. Chu XL, Zheng BS and Dai DY (1999). Fauna Sinica, Class Teleostei, Siluriformes. Scientific Press, Beijing (in Chinese).
26. Cohen NJ, Deeds JR, Wong ES, Hanner RH, Yancy HF, et al. (2009). Public health response to puffer fish (Tetrodotoxin) poisoning from mislabeled product. Journal of Food Protection 72: 810-817.
27. Cox AJ and Hebert PDN (2001). Colonization, extinction and phylogeographic patterning in a freshwater crustacean. Molecular Ecology 10: 371-386.
28. Crooijmans RPMA, Poel JJV, Groenen MMA, Bierbooms VAF and Komen J (1997). Microsatellite markers in common carp (*Cyprinus carpio*) Animal Genetics 28: 129-134.
29. Davidson WS (1998). PCR techniques. In PR Ashurst and MJ Dennis (eds.). Analytical methods of food authentication. Blackie Academic and Professional, London. 182-203 pp.
30. Dawnay NR, Ogden R, McEwing R, Carvalho GR and Thorpe RS (2007). Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Science International. Forensic Science International 173 (1):1-6.
31. Desalle R and Amato GA (2004). The expansion of conservation genetics. Nature Reviews Genetics 5:702-712.
32. Dooley JJ and Garrett SD (2001). Development of meat speciation assays using the Agilent 2100 Bioanalyzer. Agilent Technologies Application Note: 5988-4069.
33. Dooley JJ, Sage HD, Brown HM and Garrett SD (2005). Improved fish species identification by use of Lab-on-chip technology. Food Control 16: 601-607.
34. Duncan JR and Lockwood JL (2001). Extinction in a field of bullets a search for causes in the decline of the world's freshwater fishes. Biological Conservation 102: 97-105.

35. Echelle AA (1991). Conservation genetics and genic diversity in freshwater fishes of western North America. In WL Minckley and JE Deacon (eds.). Battle against extinction: native fish management in the American west. The University of Arizona Press, Tucson.
36. Englbrecht CC, Schliewen U and Tautz D (2002). The impact of stocking on the genetic integrity of Arctic charr (*Salvelinus*) populations from the Alpine region. Molecular Ecology 11:1017–1027. doi: 10.1046/j.1365-294X.2002.01498.x
37. Eschmeyer WN, Ferraris CJ Jr., Hoang MD and Long DJ (2004). The Catalog of Fishes, online, Species of Fishes. Available: http://www.calacademy.org/research/ichthyology/catalog/intro.html. Accessed July 20 2010.
38. Ferri E, Barbuto M, Bain O, Galimberti A, Uni S, et al. (2009). Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). Frontiers in Zoology 6: 1.
39. Flanders J, Jones G, Benda P et al. (2009). Phylogeography of the greater horseshoe bat, *Rhinolophus ferrumequinum*: contrasting results from mitochondrial and microsatellite data. Molecular Ecology 18: 306-318.
40. Folmer O, Black M, Hoeh W, Lutz R and Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294-299.
41. Fajardo V, Gonzalez I, Dooley J, Garret S, Brown HM, Garcia T, et al. (2009). Application of polymerase chain reaction-restriction fragment length polymorphism analysis and lab-on-a-chip capillary electrophoresis for the specific identification of game and domestic meats. Journal of the Science of Food and Agriculture 89: 843-847.
42. Ferraris Jr. CJ and de Pinna MCC (1999). Higher-level Names for Catfishes (Actinopterygii: Ostariophysi: Siluriformes). Proceedings of the California Academy of Sciences 51: 1-17.
43. FAO (2002). The State of World Fisheries and Aquaculture Fishery Resources Division, Fisheries Department, United Nations Food and Agriculture Organization, Rome. Available: http://www.fao.org/DOCREP/003/X8002E/X8002E00.htm. Accessed 23 July 2010.
44. FAO (2006). The state of world fisheries and aquaculture. Available: ftp://ftp.fao.org/docrep/fao/009/a0699e/a0699e.pdf. Accessed 20 July 2010.
45. Ferguson A, Taggart JB, Prodohl PA, McMeel O, Thompson C, Stone C, McGinnity P and Hynes RA (1995). The application of molecular markers to the study and conservation of fish populations with special reference to Salmo. Journal of Fish Biology 47(1):103-126.
46. Fausch KD, Torgersen CE, Baxter CC and Li HW (2002). Landscapes to riverscapes: Bridging the gap between research and conservation of stream fishes. Bioscience 52: 483-498.
47. Gonzales FG, Nonnis MF, Bouza C, Martinez P, Ajmone MP and Gandolfi G (2008). Application of amplified fragment length polymorphisms in gynogenetic haploid embryos of turbot (*Scophthalmus maximus*). Aquaculture and Research 39: 41-49.
48. Graham JB (1997). Air-Breathing Fishes. Evolution, Diversity and Adaptation. San Diego, CA: Academic Press.
49. Greenwood PH (1961). A revision of the genus Dinotopterus Blgr. (Pisces, Clariidae) with notes on the comparative anatomy of the suprabranchial organs in the Clariidae. Bulletin of the British Museum (Natural History) Zoology 7: 215-241.
50. Greig TW, Moore MK, Woodley CM and Quattro JM (2005). Mitochondrial gene sequences useful for species identification of western North Atlantic Ocean sharks. Fish Bulletin 103: 516-523.

51. Godinho FN and Ferreira MT (2000). Composition of endemic fish assemblages in relation to exotic species and river regulation in a temperate stream. Biology Invasions II: 231-244.
52. Gwinner HR (1973). A study of the redeye bass, *Micropterus coosae,* smallmouth bass *Micropterus dolomieu* and rockbass, *Ambloplites rupestris* in Spring Creek, Tennessee. Unpublished master's thesis. Tennessee Technological University, Cookeville, Tennessee. 37 pp.
53. Haag WR, Butler RS and Hartfield PD (1995). An extraordinary reproductive strategy in freshwater bivalves: prey mimicry to facilitate larval dispersal. Freshwater Biology 34: 471-476.
54. Hajibabaei M, Smith AM, Janzen DH, Rodriguez JJ, Whitfield JB and Hebert PDN (2006). A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6: 959-964.
55. Hajibabaei M, Singer GAC, Clare EL and Hebert PDN (2007). Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biology 5: 24 doi: 10.1186/1741-7007-5-24.
56. Hardman M (2005). The phylogenetic relationships among non-diplomystid catfishes as inferred from mitochondrial cytochrome b sequences; the search for the Ictalurid sister taxon (Otophysi: Siluriformes). Molecular Phylogenetics and Evolution 37: 700-720.
57. Hanson T (2009). Catfish database from USDA National Agricultural Statistics Service (NASS) Agricultural Statistics Board, U.S. Department of Agriculture.
58. Hebert PDN, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B 270: 313-321.
59. Hebert PD and Gregory TR (2005). The promise of DNA barcoding for taxonomy. System Biolgy 54: 852-859.
60. Heyden SVD, Barendse J, Seebregts AJ and Matthee CA (2010). Misleading the masses: detection of mislabelled and substituted frozen fish products in South Africa. ICES Journal of Marine Science 67: 176-185.
61. Hoelzel AR (2001). Shark fishing in fin soup. Conservation Genetics 2: 69-72.
62. Hogan ZS and Bernie PM (2002). Twenty-seven new microsatellites for the migratory Asian catfish family Pangasiidae. Molecular Ecology Notes 2: 38-41.
63. Hold GL, Russell VJ, Pryde SE, Rehbein H, Quinteiro J, et al. (2001). Validation of a PCR-RFLP based method for the identification of salmon species in food products. European Food Research and Technology 212: 385-389.
64. Howard JH and Morgan RP (1993). Allozyme variation in the mottled sculpin (*Cottus bairdi*): A test of stream captures hypotheses. American Society of Ichthyologists and Herpetologist. Copeia 3: 870-875.
65. Hsieh YHP, Woodward BB and Blanco AW (1995). Species substitution of retail snapper fillets. Journal of Food Quality 18: 131-140.
66. Hsieh HS, Chai T and Hwang DF (2007). Using the PCR-RFLP method to identify the species of different processed products of billfish meats. Food Control 18: 369-374.
67. Hubbs CL and Bailey RM (1940). A revision of the black basses (*Micropterus* and Huro) with descriptions of four new forms. Michigan Publication Museum of Zoology, University of Michigan. 48-51 pp.
68. Hubert N, Torrico JP, Bonhomme F and Renno JF (2008). Species polyphyly and mtDNA introgression among three Serrasalmus sister-species. Molecular Phylogenetics and Evolution 46: 375-381.

69. Hughes JM (2007). Constraints on recovery: using molecular methods to study connectivity of aquatic biota in rivers and streams. Freshwater Biology 52: 616-631.
70. Hurst H, Bass G and Hubbs C (1975). The biology of the Guadulape, Suwannee, and redeye basses. In H Clepper (ed). Black bass biology and management. Sport Fishing Institute, Washington D.C. 47-53 pp.
71. Hutchinson WF, Van Oosterhout C, Rogers SI and Carvalho GR (2003). Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society London B 270: 2125-2132.
72. Jacquet JL and Pauly D (2008). Trade secrets: renaming and mislabeling of seafood. Marine Policy 32: 309-318.
73. Kawaka M, Horbanczuk JO, Sacharczuk M, Zieba G, Lukaszewicz M, Jaszczak K and Parada R (2007). Genetic characteristics of the ostrich population using molecular methods. Poultry Science 86: 277-281.
74. Knowlton N and Weigt LA (1998). New dates and new rates for divergence across the Isthmus of Panama. Proceeding of the Royal Society B 265: 2257-2263. doi 10.1098/rspb.1998.0568.
75. Kochzius M, So¨ller R, Khalaf MA and Blohm D (2003). Molecular phylogeny of the lionfish genera Dendrochirus and Pterois (Scorpaenidae, Pteroinae) based on mitochondrial DNA sequences. Molecular Phylogenetic and Evolution 28: 396-403.
76. Kochzius M (2009). Trends in fishery genetics. In Beamish RJ and BJ Rothschild (eds). The future of fisheries science in North America. The Netherlands: Fish and Fisheries. Springer (31): 453-493.
77. Koppelman J and Garrett G (2002). Distribution, biology and conservation of the rare black bass species. In D Phillip and M Ridgway (ed). Black bass: ecology, conservation and management. American Fisheries Society 31 Bethesda, Maryland.
78. Kurtzman CP (1994). Molecular taxonomy of the yeasts. Yeast 10:1727-1740.
79. Lemer S, Aurelle D, Vigliola L, Durand J-D and Borsa P (2007). Cytochrome b barcoding, molecular systematic and geographic differentiation in rabbitfishes (Siganidae). Comptes Rendus Biologies 330: 86-94.
80. Lin WF and Hwang DF (2007). Application of PCR-RFLP analysis on species identification of canned tuna. Food Control 18: 1050-1057.
81. Lockley AK and Bardsley RG (2000). DNA-based methods for food authentication. Trends Food Science and Technology 11: 67-77.
82. Lowenstein JH, Amato G and Kolokotronis SO (2009). The real maccoyii: identifying tuna sushi with DNA barcodes-contrasting characteristic attributes and genetic distances. PLoS One 4: e7866.
83. Lowenstein JH, Burger J, Jeitner CW, Amato G, Kolokotronis SO, et al. (2010). DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. Biology Letter 6 (5): 692-695.
84. Lynch MA and Jarell PE (1993). A method for calibrating molecular clocks and its application to animal mitochondrial DNA. Genetics 135: 1197-1208.
85. MacCrimmon HR and Robbins WH (1975). Distribution of the black basses in North America. In H Clepper (ed). Black bass biology and management. Sport Fishing Institute, Washington D.C. 56-66 pp.

86. Maldini M, Marzano NF, Fortes G, Papa R and Gandolfi G (2006). Fish and seafood traceability based on AFLP markers: elaboration of a species database. Aquaculture 261: 487-494.
87. Marchetti MP and Moyle PB (2001). Effects of flow regime on fish assemblages in a regulated California stream. Ecology Application11: 530-539.
88. Mardulyn P and Whitfield JB (1999). Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids. Molecular Phylogenetic and Evolution 12: 282-294.
89. Marko PB, Lee SC, Rice AM, Gramling JM, Fitzhenry TM, et al. (2004). Fisheries: mislabelling of a depleted reef fish. Nature 430: 309-310.
90. Matthes H (1964). Les poissons du Lac Tumba et la re´gion d'Ikela. Royal Museum for Central America 8:1-199.
91. McClelland EK and Naish KA (2007). What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions.Conservation Genetics 8: 397-416.
92. Meffe GK and Vrijenhoek RC (1988). Conservation genetics in the management of desert fishes. Conservation Biology 2: 157-169. doi: 10.1111/j.1523-1739.1988.tb00167.x
93. Meyer R, Ho¨ felein C, Lu¨ thy J, and Candrian U (1995). Polymerase chain reaction-restriction fragment length polymorphism analysis: a simple method for species identification in food. Journal of the AOAC International 78: 1542-1551.
94. Miller DD and Mariani S (2010). Smoke, mirrors, and mislabeled cod: poor transparency in the European seafood industry. Frontiers in Ecology and the Environment .Available: http://www.esajournals.org/doi/abs/10.1890/090212 Accessed 15 July 2010.
95. Mohindra V, Singh RK, Palanichamy M, Ponniah AG and Lal KK (2007). Genetic identification of three species of the genus Clarias using allozyme and mitochondrial DNA markers. Journal of Applied Ichthyology 23: 104-109.
96. Moore WS (1995). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49: 718-726.
97. Moritz C (1999). Conservation units and translocations: strategies for conserving evolutionary processes. Hereditas 130: 217-228.
98. Mount JF (1995). California Rivers and Streams: The Conflict between Fluvial Process and Landuse, University of California Press, Berkeley. 359 pp.
99. Na-Nakorn U (1999). Genetic factors in fish production: a case study of the catfish *Clarias*. In S Mustafa (ed). Genetics in Sustainable Fisheries Management. London: Fishing News Books.175-187pp.
100. Nanney DL (1982). Genes and phenes in Tetrahymena. Bioscience 32: 783-788.
101. Nelson K and Soule M (1987). Genetical conservation of exploited fishes. In N Ryman and F Utter (eds). Population genetics and fishery management. University of Washington Press, Seattle. 345-371pp.
102. Nielsen R and Matz M (2006). Statistical approaches for DNA barcoding. System Biology 55:162-169.
103. Nielsen EE and Hansen MM (2008). Waking the dead: the value of population genetic analyses of historical samples. Fish and Fisheries 9: 450-461.

104. Ochiai Y, Ochiai L, Hashimoto K and Watabe S (2001). Quantitative estimation of dark muscle content in the mackerel meat paste and its products using antisera against myosin light chains. Journal of Food Science 66:1301-1305.
105. Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276: 734-740.
106. Panaro NJ, Yuen PK, Sakazume T, Fortina P, Kricka LJ and Wilding P (2000). Evaluation of DNA fragments sizing and quantification by the Agilent 2100 Bioanalyzer. Clinical Chemistry 46:1851-1853.
107. Parson JW (1954). Growth and habits of the redeye bass. Transaction of American Fisheries Society 83: 202-211.
108. Philipp D (1991). Genetic implications of introducing Florida largemouth bass, *Micropterus salmoides floridanus*. Canadian Journal of Fisheries and Aquatic Sciences 48: 58-65.
109. Pipas J and Bulow F (1998). Hybridization between redeye bass and smallmouth bass in Tennessee streams. Transactions of the American Fisheries Society 127: 141-146.
110. Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE and Stromberg JC (1997). The natural flow regime: Aparadigm for river conservation and restoration. Bioscience 47: 769-784.
111. Pouyard L, Teugels GG, Gustiano R and Legendre M (2000). Contribution of the phylogeny of *Pangasius* catfishes based on allozymes and mitochondrial DNA. Journal of Fish Biology 56: 1509-1538.
112. Proteau JP, Hilge V and Linhart O (1996). État actuel et perspectives de la production aquacole des poissons-chats (Siluroidei) en EuropePresent state and prospects of the aquaculture of catfishes (Siluroidei) in Europe. Aquatic Living Resources 9 (Special Issue): 229-235.
113. Quinteiro J, Sotelo CG, Rehbein H, Pryde SE, Medina I, et al. (1998). Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR restriction fragment length polymorphism methodologies in species identification of canned tuna. J Agric Food Chem 46: 1662-1669.
114. Rasmussen RS, Morrissey MT and Hebert PD (2009). DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. Journal of Agriculture and Food Chemistry 57: 8379-8385.
115. Ratnasingham S and Hebert PD (2007). BOLD: The Barcode of Life Data System. Molecular and Ecology Notes 7: 355-364.
116. Rhymer J and Simberloff D (1996). Extinction by hybridization and introgression. Annual Review in Ecology and Systematics 27: 83-109.
117. Rohde F, Arndt R, Lindquist D and Parnell J (1994). Freshwater Fishes of the Carolinas, Virginia, Maryland, and Delaware. The University of North Carolina Press, Chapel Hill, North Carolina.
118. Roheim C and Sutinen J (2006). Trade and Marketplace Measures to Promote Sustainable Fishing Practices. ICTSD Natural Resources, International Trade and Sustainable Development Series Issue Paper 3. International Centre for Trade and Sustainable Development and the High Seas Task Force.Geneva. 48 p.
119. Ross R (1971). The drainage history of the Tennessee River. In P Holt (ed). The Distributional History of the Biota of the Southern Appalachians Part III: Vertebrates.

Research Division Monograph 4. Virginia Polytechnic Institute and State University, Blacksburg, VA.
120. Rubidge E and Taylor E (2004). Hybrid zone structure and the potential role of selection in hybridizing populations of introduced rainbow trout (*Oncorhynchus mykiss*) and native westslope cutthroat trout (*O. clarki lewisi*). Molecular Ecology 13: 3735-3749.
121. Ruzzante DE, Taggart CT and Cook D (1998). A nuclear DNA basis for shelf- and bank-scale population structure in northwest Atlantic cod (*Gadus morhua*): Labrador to Georges Bank. Molecular Ecology 83:1663-1680.
122. Salini JP, Milton DA, Rahaman MJ and Hussein MG (2004). Allozyme and morphological variation throughout the geographic range of the tropical shad, hilsa *Tenualosa ilisha*. Fisheries Research 66: 53-69.
123. Sangthong P and Jondeung A (2003). Cloning and nucleotide sequence of four tRNA genes in mitochondrial genome of Thai walking catfish (*Clarias macrocephalus*) Kasetsart Journal Natural Science 37:33-40.
124. Sanjuan A and Comesana AS (2002). Molecular identification of nine commercial flaffish species by polymerase chain reaction-restriction fragment length polymorphism analysis of a segment of the cytochrome b region. Journal of Food Protection 65: 1016-1023.
125. Santaclara FJ, Cabado AG and Vieites JM (2006). Development of a method for genetic identification of four species of anchovies: *E. encrasicolus, E. anchoita, E. ringens* and *E. japonicus*. European Food and Research Technology 223: 609-614.
126. Shaw PW, Turan C, Wright JM, O'Connell M and Carvalho GR (1999). Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. Heredity 83: 490-499.
127. Simmons RB and Weller SJ (2001). Utility and evolution of cytochrome b in insects. Molecular Phylogenetic and Evolution 20: 196-210.
128. Smith PJ, Roberts CD, McVeagh SM and Benson PG (1996). Genetic evidence for two species of tarakihi. Teleostei: Cheilodactylidae: *Nemadactylus* in New Zealand waters. New Zealand Journal of Marine and Freshwater Research 30: 209-220.
129. Smith PJ, Veagh SMM and Steinke D (2008). DNA barcoding for the identification of smoked fish products. Journal of Fish Biology 72: 464-471.
130. Sotelo CG, Pineiro C, Gallardo JM and Perez-Martin RI (1992). Identification of fish species in smoked fish products by electrophoresis and isoelectric focusing. Zeitschrift für Lebensmittel-Untersuchung und -Forschung 195: 224-227.
131. Sotelo CG, Calo-Mata P, Chapela MJ, Pe´rez-Martı´n RI, Rehbein H, et al. (2001). Identification of flatfish (Pleuronectiformes) species using DNA-based techniques. Journal of Agriculture and Food Chemistry 49: 4562-4569.
132. Steinke D, Vences M, Salzburger W and Meyer A (2005). TaxI: a software tool for DNA barcoding using distance methods. Philosophical Transactions of the Royal Society of London B 360: 1975-1980.
133. Steinke D, Zemlak TS and Hebert PD (2009). Barcoding nemo: DNA-based identifications for the ornamental fish trade. PLoS One 4 (7): e6300.
134. Strange EM, Moyle PB and Foin TC (1992). Interactions between stochastic and deterministic processes in stream fish community assembly. Environmental Biology Fishery 36: 1-15.
135. Swift C, Gilbert C, Bortone S, Burgess G and Yerger R (1986). Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Pontchartrain.

In C Hocutt and E Wiley (ed). The zoogeography of North American freshwater fishes. John Wiley and Sons New York, New York.
136. Tatum WR (1965). A study of an introduced population of redeye bass, *Micropterus coosae* Hubbs and Bailey, in Spring Creek, Tennessesee. Unpublished master's thesis, Tennessee Technological Univeristy, Cookville, Tennessee. 35pp.
137. Teletchea F (2009). Molecular identification methods of fish species: reassessment and possible applications. Review in Fish Biology and Fisheries 19: 265-293.
138. Tepedino V, Berrini A, Borromeo V, Gaggioli D, Cantoni C, Manzoni P and Secchi C (2001). Identification of commercial fish species belonging to the orders Pleuronectiformes and Gadiformes: library of isoelectric focusing patterns. Journal of AOAC International, 84(5): 1600-1607.
139. Teugels GG, Sudarto L and Pouyard L (2001). Description of a new *Clarias* species from Southeast Asia based on morphological and genetical evidence (Siluriformes, Clariidae). Cybium 25: 81-92.
140. Teugels GG and Adriaens D (2003). Taxonomy and phylogeny of Clariidae - an overview. In G Arratia, BG Kapoor, M Chardon and R Diogo (ed). Catfishes. Enfield, News Hamsphire, USA: Science Publishers: 465-487.
141. Thuesen PA, Pusey BJ, Peck DR, Pearson R and Congdon BC (2008). Genetic differentiation over small spatial scales in the absence of physical barriers in an Australian rainforest stream fish. Journal of Fish Biology 72: 1174-1187. doi 10.1111/j.1095-8649.2007.01768.x.
142. Unlusayin M, Kaleli S and Gulyavuz H (2001). The determination of flesh productivity and protein components of some fish species after hot smoking. Journal of the Science of Food and Agriculture 81: 661-664.
143. USDA (2002). Catfish Production. National Agricultural Statistics Service (NASS), Agricultural Statistics Board, U.S. Department of Agriculture. Available: http://www.usda.gov/nass Accessed 2 Aug 2010.
144. U.S Fish and Wildlife Service (1994). Endangered and threatened wildlife and plants, 50 CFR 17.11 and 17.12. Division of Endangered Species, US Fish and Wildlife Service, Washington D.C. 42 pp.
145. Uthayakumaran S, Batey IL and Wrigley CW (2005). On-the-spot identification of grain variety and wheat-quality type by Lab-on-a-chip capillary electrophoresis. Journal of Cereal Science 41: 371-374.
146. Vasilyeva E, Woodard J, Taylor FR, Kretschmer M, Fajardo H, Lyubarskaya Y, Kobayashi K, Dingley A and Mhatre R (2004). Development of a chip-based capillary gel electrophoresis method for quantification of a half-antibody in immunoglobulin G4 samples. Electrophoresis 25(21-22): 3890-3896.
147. Ward JV and Stanford JA (1983). The serial discontinuity concept of lotic ecosystems. In TD Fontaine and SM Bartel (eds). Dynamics of Lotic Ecosystems. Ann Arbor Science Publications, MI. 29-42pp.
148. Ward JV, Tockner K, Uehlinger U and Malard F (2001). Understanding natural patterns and processes in river corridors and the basis for effective river restoration. Regulation Rivers Restoration Management 17: 311-323.
149. Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PD (2005). DNA barcoding Australia's fish species. Philosophical Transaction of Royal Society B. Biology Science 360: 1847-1857.

150. Ward RD, Hanner R and Hebert PDN (2009). The campain to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329-356.
151. Wares JP and Cunningham CW (2001). Phylogeography and historical ecology of the North Atlantic Intertidal. Evolution 12: 2455-2469.
152. Watanabe T, Fujita H, Yamasaki K, Seki S and Taniguchi N (2004). Preliminary study on linkage mapping based on microsatellite DNA and AFLP markers using homozygous clonal fish in ayu (*Plecoglossus altivelis*). Marine Biotechnology 6: 327-334.
153. Wilson KH (1995). Molecular biology as a tool for taxonomy. Clinical Infectious Diseases 20: 192-208.
154. Wofford JEB, Gresswell RE and Banks MA (2005). Influence of barriers to movement on within-watershed genetic variation of coastal cutthroat trout. Ecological Applications 15: 628-637.
155. Wong EHK and Hanner RH (2008). DNA barcoding detects market substitution in North American seafood. Food Research International 41: 828-837.
156. Yamamoto S, Morita K, Koizumi I and Maekawa K (2004). Genetic differentiation of white-spotted charr (*Salvelinus leucomaenis*) populations after habitat fragmentation: spatial-temporal changes in gene frequencies. Conservation Genetics 5: 529-538.
157. Yeager BL and Saylor CF (1995). Fish hosts for four species of freshwater mussel (Pelecypoda: Unionidae) in the upper Tennessee River drainage, American Midland Naturalist 133: 1-6.
158. Yue PQ and Chen YY (1998). China Red Data Book of Endangered Animals: Pisces. Beijing: Science Press. 247p.
159. Zale AV and Neves RJ (1982). Fish hosts of four species of lampsiline mussels (Mollusca: Unionidae) in Big Moccassin Creek, Virginia. Canadian Journal of Zoology 60: 2535-2542.
160. Zhang DX and Hewitt GM (1997). Assessment of the universality and utility of a set of conserved mitochondrial primers in insects. Insect Molecular Biology 6:143-150.