NOTE TO USERS

This reproduction is the best copy available.

UMI®

PHYLOGEOGRAPHY AND CONSERVATION GENETICS OF SCALELESS CARP (*GYMNOCYPRIS PRZEWALSKII*) OF LAKE QINGHAI, CHINA

by

Dana M. O'Bryan

B.Sc. (Hon.), University of British Columbia, 1999

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OFF
MASTER OF SCIENCE
IN
NATURAL RESOURCE AND ENVIRONMENTAL STUDIES

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

April 2010

Your file Votre référence
ISBN: 978-0-494-60867-8
Our file Notre référence
ISBN: 978-0-494-60867-8

ABSTRACT

The threatened scaleless carp, *Gymnocypris przewalskii*, is endemic to saline Lake Qinghai. This study used genetic techniques to examine the phylogeography of scaleless carp, specifically to look for the presence of a distinct stock structure. Significant differences in mitochondrial DNA variation were found between scaleless carp, its ancestor, *Gymnocypris eckloni*, and its subspecies, *G. p. ganzihonensis*. Amplified Fragment Length Polymorphism analysis found significant differences between most fish sampled in tributaries, indicating that spawning groups should be treated as Management Units for conservation purposes.

Analysis of Major Histocompatibility Complex (*Mhc*) variation in scaleless carp identified two locus groups, and found that these genes are highly polymorphic, and in terms of function and selection, behave in a similar fashion as in other teleosts. Additional information is needed if *Mhc* is to be fully characterized and used to further assess the genetic differences between *G. p. przewalskii, G. eckloni,* and *G. p. ganzihonensis.*

ACKNOWLEDGEMENTS

I would like to first offer my thanks to all of my fellow colleagues in "Lab 4-403" for their constant technical assistance and support. I would especially like to thank Angie Dale, who helped me immensely with the AFLP methodology and analysis.

I would like to offer my gratitude to my supervisory committee, Dr. Mark Shrimpton and Dr. Andrea Gorrell, for their guidance.

This thesis would not have been possible without the constant encouragement and support (financial and otherwise) of my supervisor, Dr. Brent Murray. Many thanks for your patience and never allowing me to give up, and for allowing me to participate in the research trip to China. Who knew that the best homemade potato chips in the world would come from Xining?

I must thank my husband, Eric, for supporting me throughout my time as a graduate student, and for never allowing me to utter the word "quit".

And finally, I must give my thanks to my beautiful twin boys, Hayden and Liam, for their extra long afternoon naps and early bedtimes, without which I never would have had the time to write. And for making my life as much about finger painting and play dough as it was about AMOVAs and phylogenies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COMMON ABBREVIATIONS

AFLP: Amplified Fragment Length Polymorphism

AMOVA: Analysis of Molecular Variance

bp: base pair

ESU: Evolutionary Significant Unit

Mhc: Major histocompatibility complex

mtDNA: Mitochondrial DNA

MU: Management Unit

NE Units: New England Biolabs Units

1 GENERAL INTRODUCTION

1.1 LAKE QINGHAI AND THE QINGHAI-TIBET PLATEAU

The Qinghai-Tibet Plateau (Figure 1.1), located in western China and described as the "Roof of Asia", is an immense fault block formed by tectonic movements in the Tertiary, when the Indian and Eurasian plates converged (Walker *et al.*, 1996). The subsequent uplift of the Himalayan, Karakorum, and Kunlun mountains blocked the path of monsoons from the south, which led to increasing aridity in the region. After the formation of the plateau, drainage lines to the Indian Ocean were blocked, forming numerous lakes - over 1600 small lakes, which may be ephemeral depending on precipitation and proximity to rivers, and 10 large lakes, together covering an area over 40,000 km^2 (Walker and Yang, 1999). Those that were located in terminal basins became progressively more saline; hence, the region contains approximately 350 saltwater lakes, which range in salinity from slightly brackish to crystallising brines.

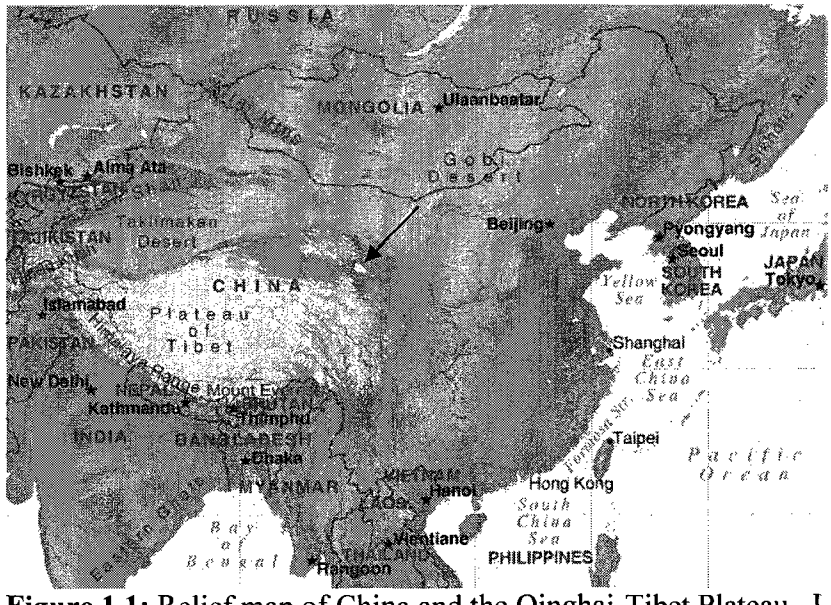


Figure 1.1: Relief map of China and the Qinghai-Tibet Plateau. Lake Qinghai is indicated by an arrow (cf. http://www.map-of-china.org/china-relief-map.gif).

Lake Qinghai is located within the Chaidmu Basin of the plateau (along the eastern edge, making it 'the gateway to the plateau') at an altitude of 3200 m. The ancestral lake was formed approximately 500,000 years ago as a result of the downfaulting of the Tibetan Plateau (Bian *et al.*, 2000). This lake was glacier-fed and originally drained into the Yellow River; however, its connection was severed approximately 130,000 to 150,000 years ago after an uplift of the plateau (termed the Gonghe Movement). This turned the entire Lake Qinghai drainage inward, forming the endorheic basin that comprises the lake today (Zheng, 1997; Li and Fang, 1998). Since the lake is located in a dry, desert area, its net inflow is low and its evaporation is high. The lake has regressed to about one-third of its original size, becoming saline in the process. It is currently the largest saline lake in China (surface area 4600 km^2) with a high pH (9.4) and a salinity of 13.13 g/L (Wood *et al.*, 2007). Several freshwater rivers drain into the lake, the largest of which is the Buha River, which, when combined with the Shaliu and Haerga Rivers, contributes about 80% of the total annual inflow (Walker and Yang, 1999). Smaller drainages include the Heima and Quanji rivers, and formerly the Ganzi River, which is no longer directly linked with the lake (see Sections 1.2 and 2.2). The lake also contains seasonal water bird sanctuaries, Bird Island and Egg Island, which support over 100,000 bar-headed geese, black-headed gulls, and cormorants in a breading area of 67,000 m^2. Incidentally, of the five total islands located on Lake Qinghai, three, including Bird and Egg Islands, became peninsulas in the 1970s as a result of falling lake levels and siltation from rivers (Walker *et al.*, 1996).

1.2 SCALELESS CARP AND OTHER FISH SPECIES OF THE TIBETAN PLATEAU

The Qinghai-Tibetan Plateau supports a total of 112 native and 17 introduced fish species. The native species are believed to be the relicts of mass extinction events that occurred in the Quartenary, during the uplift of the Himalayan Mountains (Walker and Yang, 1999).

Although Lake Qinghai is the largest in China, due to its extreme environment (low temperature, high pH, saline, low oxygen), it has few resident fish species, the most abundant of which is the naked, or scaleless carp, *Gymnocypris przewalskii* (Cyprinidae: Schizothoracinae). The Schizothoracine subfamily, which includes all naked carp or "snow trout", is an evolutionary offshoot of the Cyprinidae, probably arising from a surviving Barbinae (Walker and Yang, 1999). It occurs throughout central and southern Asia (Banarescu and Coad, 1991), although more than 80% of species are located in China. The subfamily includes 11 genera in three phylogenetic groups (Wu, 1984). *Gymnocypris* is a member of the third, specialized group, which tends to occur mainly at higher altitudes. While all fish from the subfamily share common traits, including slow growth and general lack of scales (some have small columns of scales around the anus), only *G. przewalskii* has been studied in any detail, possibly because it is the species with the most commercial value in the region.

The scaleless carp commercial fishery began in 1950s, with the total catches peaking in 1960 (a record catch of more than 28,000 tonnes). Catches then declined rapidly, as did the average length and weight of fish, following the typical pattern of newly exploited fisheries (Walker and Yang, 1999). A total of 200,000 t was produced by the Lake Qinghai fishery, from 1958 until 1991, when a moratorium was put in place. The fish are very slow growing (a result of the cold climate), taking approximately 7 years to reach sexual maturity and a marketable weight of 300 g; therefore, stocks have not recovered from overexploitation (Walker *et al.*, 1996). Poaching of the species continues today and the commercial fishery may be reopened in the future.

A 1987 census showed that the region contained 90,500 people, most who work in agriculture (Walker *et al.*, 1996). Although fisheries have been established in western China for

more than 40 years, they contribute a very small portion of total agricultural production, partially due to slow markets and poor economic conditions. In the Lake Qinghai area in particular, the local people tend to eat mostly mutton and beef, and Tibetans are generally reluctant to eat fish for religious reasons. However, a rapidly expanding population in the Lake Qinghai area (and in other areas of the Tibetan Plateau), especially of the Han people from other areas of China, have encouraged the expansion of capture fisheries (Walker *et al.*, 1996; Walker and Yang, 1999).

Other species of the *Gymnocypris* genera are fished commercially, including *G. eckloni*, from the Yellow River and Gyaring and Ngoring Lakes to the south of Lake Qinghai, and *G. waddelli*, located in Yamdrok Lake in Tibet (Table 1.1). Although their catches are much smaller when compared to historical catches of *G. przewalskii*, catches may also increase due to a growing human population. Other *Gymnocypris* species of the Qinghai-Tibetan Plateau are included in Table 1.1.

Table 1.1: *Gymnocypris* species of the Qinghai-Tibet Plateau

Species	**Location**	**Salinity (g/L)**	**pH**
G. chui	Lake Langchu	0.941	8.6
G. dobula	Lake Peigucco	1.921	9.5
G. eckloni	Yellow River, Lakes Gyaring and Ngoring	Freshwater	
G. namensis	Lake Nam Co	1.715	9.3
G. potanii	Mekong and Yangtze Rivers	Freshwater	
G. przewalskii	Lake Qinghai	13.13	9.3
G. waddellii	Lake Yamdrok	1.781	9.3

*Information in this table was taken from a variety of sources: Walker *et al.* (1996), www.fishbase.org, and Yuan (1990).

Based on morphological and electrophoretic characters, a subspecies of the scaleless carp, *Gymnocypris przewalskii ganzihonenesis*, has been identified in the Ganzi River, which at one time was connected to Lake Qinghai. It is believed that the connection to the lake was severed, isolating fish populations residing within the river and leading to the evolution of a possible subspecies (Walker and Yang, 1999). However, the status of the Ganzi River fish population is unknown as there have been no reports of them being examined since they were first described (Zhu and Wu, 1975).

1.3 UNUSUAL PHYSIOLOGY OF SCALELESS CARP

Water levels in Lake Qinghai have fallen approximately 12 m in the last hundred years, due to a combination of decreased rainfall, increased evaporation, and the diversion of water in tributaries for agriculture. This has resulted in elevated salinity levels, as well as the formation of two additional lakes (Erhai and Gahai) along the eastern edge (Walker and Yang, 1999). Remarkably, scaleless carp are currently able to tolerate the high salinity by allowing most plasma ions to equilibrate with the lake water, thereby reducing the energy demands required for ion regulation (Wood *et al.*, 2007). This ability to thrive in saline waters has made scaleless carp a candidate for aquaculture expansion in other salt lakes of the plateau, which have few or no fish species (Walker *et al.*, 1996). However, the increasing salinity of Lake Qinghai may soon have detrimental physiological effects on resident fish populations. Studies have indicated that the waters of Lake Qinghai are very near the salinity threshold for scaleless carp, as even small increases in salinity (4 ppt) resulted in osmoregulatory disturbances that resulted in death or severe loss of equilibrium (Brauner *et al.*, 2008).

Unlike other cyprinids, scaleless carp undergo an annual spawning migration between March and July from the saline lake to adjoining freshwater rivers (Wang *et al.*, 2003). The

transfer between the lake water and river water poses additional physiological challenges to the fish in terms of nitrogenous waste excretion and ion balance. However, once again, the scaleless carp uniquely adapt to this migration by increasing plasma Na^+ and Cl^- ions and lowering K^+ ions (Wang *et al.*, 2003). Most euryhaline teleosts lower Na^+ and Cl^- when transferred to freshwater from brackish water due to salt loss via passive diffusion (Perry and Laurent, 1989; Postlethwaite and McDonald, 1995). It is thought the higher levels of plasma ions allow the fish to meet the energetic demands of up-river migration (Wang *et al.*, 2003). The scaleless carp are also well adapted to the chronically reduced oxygen levels of the lake (6 mg O_2/l), and are uniquely tolerant of long-term low oxygen conditions (more so than its hypoxia-tolerant relative, the crucian carp) (Matey *et al.*, 2008).

1.4 GENERAL RESEARCH OBJECTIVES

In addition to the physiological challenges of increasing salinity within the lake, the installation of impassable weirs and dams on rivers and increased siltation have resulted in a vast loss of spawning habitat. In response to management and conservation concerns for fish populations, an international collaborative project between Canada (Queen's University, University of Northern British Columbia, University of British Columbia, and McMaster University) and China (Zhejiang University) was initiated to study the evolution, ecology, and physiological adaptations of the scaleless carp and related species to the environments of the Tibetan Plateau. The results of the physiological studies have been reported elsewhere (Wood *et al.*, 2007; Matey *et al.*, 2008) and further publications are pending. This thesis analyzes genetic variation in order to examine the phylogeography and systematics of scaleless carp. Specifically, this study will look for the presence of a distinct stock structure of *G. p. przewlaskii* within the lake, and will also examine the relationships between the Lake Qinghai species, its subspecies,

G. p. ganzihonensis, and its ancestral species, *G. eckloni*. The identification of different stocks within Lake Qinghai is extremely important for the conservation and management of scaleless carp populations both within the lake and in each river. Some tributaries, and associated spawning habitat, may be lost due to falling lake levels, agricultural divergence, and damming. If a tributary is lost, an entire distinct population may become extinct, posing a threat to the genetic variation of the entire Lake Qinghai species. If distinct stocks are not present, the loss of spawning habitat will still be detrimental to the species; however, usable habitat may be recreated at a later date. The Ganzi River, now isolated from the lake, has been heavily impacted by anthropogenic disturbances, and information on this isolated fish population is critical to initiation of conservation actions. Phylogeographic relationships within the Lake Qinghai system will be determined by examining the sequence variation in the control region of mtDNA, in combination with Amplified Fragment Length Polymorphisms (AFLPs) (Chapter 2).

A secondary goal of this project is to examine sequence variation in the Major Histocompatibility Complex (*Mhc*) in order to determine if it can be developed as a genetic marker system for the management of scaleless carp stocks (Chapter 3).

1.5 THE USE OF GENETIC MARKERS IN FISH CONSERVATION

The ability to differentiate distinct populations is central to management of fish stocks (Carvalho and Hauser, 1994). Stocks are often identified using tagging methods, such as coded wire tags, which depend on the recovery of the tags from dead fish. As lethal methods are not usually recommended for imperilled fish populations, such as the scaleless carp, genetic analysis may be preferred. The ideal genetic technologies for stock analysis are those based on biological variation in characters that differ substantially among stocks, show little temporal or annual variation within stocks, and can be screened in a rapid, nonlethal, and cost effective manner

(Beacham *et al.*, 2001). The choice of appropriate genetic method will depend on the level of accuracy and precision of the stock composition estimates required for management, the ease of obtaining those estimates, and the cost of the analysis. Allozymes can provide regional estimates of stock composition if there is a regional population strucuture; however, a finer scale technique may be more effective when trying to discriminate populations beyond the regional scale. Protein electrophoresis has been used in coho salmon, but there is a limited amount of variation present at protein electrophoretic loci (Wehrhahn and Powell, 1987; Bartley *et al.*, 1994). Microsatellite loci are excellent markers in stock structure and genetic population studies because they are abundant within the genome, highly polymorphic, and, since they are non-coding, selectively neutral. The use of microsatellites is a good choice for generating estimates of gene flow, effective population size, and phylogenetic relationships (Beacham *et al.*, 2001). However, start up times and costs are large, especially in organisms for which there is very little prior genetic information, like *Gymnocypris.* While the research for this thesis was in progress, microsatellites were developed for scaleless carp (Zhang *et al.*, 2009), although the size and scope of this study is much larger than the study performed using microsatellites.

This study uses the control region of mitochondrial DNA and Amplified Fragment Length Polymorphisms (AFLPs). The mtDNA control region is generally selectively neutral (non-coding) and provides information on non-selective processes, such as mutation and drift. Mitochondrial DNA is frequently used in population genetic structure studies (Brunner *et al.*, 1998; Stepien and Faber, 1998; Taylor *et al.*, 2001; Gharrett *et al.*, 2005) and to assess subspecies (Tabata and Taniguchi, 2000; April and Turgeon, 2006). Additionally, there are an increasing number of fish studies that combine mitochondrial variation with nuclear markers,

such as AFLPs (Young *et al.*, 2001; Seehausen *et al.*, 2003; Ishikawa *et al.*, 2004; Mock and Miller, 2005).

AFLPs, developed by Vos and colleagues (1995), combine restriction digestion with two levels of PCR amplification to generate multiple loci across the entire nuclear genome, rather than only examining one gene or region. AFLP analysis is especially well suited to studies of organisms with very little prior genetic information, as in *G. przewalskii,* because non-specific primers are used. AFLPs have fewer artifacts and errors than other non-specific techniques such as RAPDs (Mueller and Wolfenbarger, 1999) and are capable of identifying relationships at the species or population level, providing the resolving power to detect the early divergence of lineages (Sullivan *et al.*, 2004). Therefore, AFLPs are a suitable tool for examining the stock structure within Lake Qinghai, as well as genetic variation of *G. przewalskii ganzihonensis* in the Ganzi River and *G. eckloni* in the Yellow River (Mickett *et al.*, 2003; Whitehead *et al.*, 2003; Elderkin *et al.*, 2004; Bensch and Åkesson, 2005). AFLP analysis has previously been performed on *Gymnocypris przewalskii*; however, the goal of that study was to assess the genetic diversity of the fish rather than to detect significant differences between spawning groups (*e.g.* no inferential statistics were performed) (Chen *et al.*, 2005).

This thesis will also study genes of the Major Histocompatibility Complex (*Mhc*) (Chapter 3), which is a multi-gene family that acts as the interface between the immune system and infectious diseases (Bernatchez and Landry, 2003). There are two major classes of *Mhc* molecules that code for proteins involved in antigen presentation (Staines *et al.*, 1993; Bernatchez and Landry, 2003; Rakus *et al.*, 2003). Class I molecules are expressed on all nucleated cells and bind endogenously produced foreign peptides (usually viral) in the host cell's cytoplasm. The bound peptides are then presented on the cell surface to be recognized by

cytotoxic T cells, initiating further immune response. Class II molecules are expressed only on specific cells of the immune system (*e.g.* B cells, macrophages, etc.), and recognize and bind exogenous foreign peptides (*e.g.* as in a bacterial infection) to be presented to helper T cells.

Mhc genes have several assets that make them highly useful as molecular markers for scaleless carp populations. First, the molecular processes of the *Mhc* have been extensively studied, and there is ample information about the structure and function of the genes in a variety of teleosts, including close relatives of *Gymnocypris* (Dixon *et al.*, 1996). Second, although the general architecture of *Mhc* families is relatively conserved across species, the number of loci varies substantially. In addition, within a species, the *Mhc* is highly polymorphic, such that it is very unlikely for two unrelated individuals to have identical *Mhc* make-ups (Bernatchez and Landry, 2003). High levels of polymorphism result in the presence of numerous alleles and haplotypes within a population.

The ability of *Mhc* proteins to bind foreign peptides is partly determined by polymorphism in the amino acid residues of the Peptide Binding Region (PBR), which is thought to be maintained by balancing selection (*i.e.* heterozygote advantage and low frequency dependent selection) (Klein, 1987; Dixon *et al.*, 1996). However, recent speciation events or lineage divergence may result in greater selective pressure, which is most intense at the peptide binding region. As *Mhc* offers high polymorphism and evolves trans-specifically (shared allelic lineages across species), it is an ideal marker to assess phylogenetic relationships in closely related species and species that have evolved rapidly (Málaga-Trillo *et al.*, 1998).

Although the adaptive nature of *Mhc* genes would generally preclude their use in studies that require the assumption of selective neutrality, *Mhc* allele frequencies have the potential to

enhance stock specificity and provide quantitative information on adaptive variation in populations (Beacham *et al.*, 2004). A recent study reported that one *Mhc* locus with 15 identified alleles was more effective for stock identification than 13 microsatellites (Beacham *et al.*, 2004). The high levels of variability in *Mhc* genes also allow the differentiation of populations where other genetic markers show little or no variation (Cohen *et al.*, 2006). For the future management of scaleless carp (especially if policy changes allow the fishery to be officially re-opened), the development of a genetic database may allow the identification of mixed stock samples.

1.6 DEFINING UNITS FOR CONSERVATION

When discussing conservation practises for a threatened species or population, one must first be able to distinguish the taxa or "units" to be conserved. First, there is the species, which would seem to be relatively easy to identify; however, a study by Hey (2001) listed two dozen different "concepts" that have been proposed to describe what is a species. In fact, there is likely no greater debate in the field of biology. To list and debate these different concepts is beyond the scope of this thesis, but the most popular (and the most relevant to this thesis) are the Biological Species Concept (Dobzhansky, 1951; Mayr, 1963), where a species is a group of interbreeding natural populations that is reproductively isolated from other groups (Mayr and Ashlock, 1991), and the Phylogenetic Species Concept (Cracraft, 1983), where a species is a monophyletic group composed of the smallest diagnosable cluster of individuals within which there is a parental pattern of ancestry and descent.

Further complicating matters is the inclusion of intraspecific taxa or "subspecies". Taxonomically speaking, a subspecies is simply the taxa below the species level and is often defined by variations in morphological traits (Haig *et al.*, 2006). However, much like species

concepts, definitions of subspecies are a source of considerable debate among biologists. A 2006 review by Haig and colleagues found no universally accepted definition of a subspecies. Mayr (1963) defined a subspecies as geographically isolated from other members of its species, while many others consider subspecies to be incipient species, on the evolutionary path to complete divergence (Frankham *et al.*, 2002). While a subspecies can exist under the Biological Species Concept, it cannot under the Phylogenetic Species Concept, since species are defined as the smallest monophyletic unit, and groupings below the species level would not be considered taxonomically relevant.

The Evolutionary Significant Unit (ESU) is one or a set of conspecific populations with a distinct, long-term evolutionary history mostly separate from other such units (Ryder, 1986; Avise, 2000). The term was initially proposed to provide a rational basis for prioritizing taxa for conservation purposes with an emphasis on preserving historical population structure (Moritz, 1994). Again, there are differences of opinions as to how one recognizes an ESU since standard definitions have evolved over time. An ESU was first defined as population units that "represent significant adaptive variation based on concordance between sets of data derived by different techniques" (Ryder, 1986). ESUs were later defined as "reproductively separate from other populations" with unique adaptations, thus contributing substantially to the overall genetic diversity of a species (Waples, 1991). Moritz (1994) proposed a more quantitative approach by defining an ESU as a distinct group that exhibits reciprocal monophyly at mtDNA alleles and shows significant divergence of allele frequencies at nuclear genes (though it does not qualify how many nuclear genes or define what is considered "significant divergence").

In many cases, populations may exhibit statistically significant differences in allele frequencies, without exhibiting reciprocal monophyly at mtDNA alleles. Under Moritz's (1994)

definition, such populations would not be considered ESUs; however, they would still warrant special consideration for conservation purposes. Hence, the term "Management Unit" (MU) was proposed (Moritz, 1994). While an ESU is concerned with historical population structure and warrants long term management and protection, a MU (sometimes used interchangeably with the term 'stock') represents current population structure, warranting monitoring and short-term management (Moritz, 1994).

However, reciprocal monophyly at mtDNA alleles, which is required under Moritz's definition of an ESU, is unlikely to be found in species or populations characterized by high levels of gene flow, or those that have experienced a relatively recent speciation event (Crandall *et al.*, 2000). Defining such groups merely as MUs is insufficient; therefore, in this study, scaleless carp populations will also be examined using an alternative method - the principles of ecological and genetic exchangeability (Crandall *et al.*, 2000). Crandall and colleagues stressed that both genetic and ecological data, both recent and historic in nature, should be used in defining population distinctiveness. Ecological exchangeability refers to the ability of individuals to be moved between populations and share the same niche, which arises from shared fundamental adaptations such as life history traits, morphology, ecological requirements, and demographic characteristics, whereas, genetic exchangeability refers to high levels of gene flow between populations. Unique alleles, low gene flow estimates, and phylogenetic divergence concordant with geographic barriers would all indicate a lack of genetic exchangeability between populations (Crandall *et al.*, 2000). The ESU definition can be considered very 'black or white' (ESU or not) and tends to focus mostly on identifying genetic isolation and preserving historical legacy, while the principles of exchangeability aim to preserve adaptive diversity by recognizing species as consisting of continually evolving populations with varying levels of gene flow.

1.7 ORGANIZATION OF THESIS

This thesis has been prepared in "sandwich" format. Chapter 2 represents a manuscript[1] which has been submitted for publication to the Journal of Fish Biology, where it is currently under review. Chapter 3 has been written in the same manner, although there are no plans to submit it for publication in its current format as it is only a preliminary study and would require more data prior to submission. A brief General Discussion follows Chapter 3. All references have been compiled into a single section located after the General Discussion.

Three appendices are also included – Appendix 1 includes a mtDNA parsimony network figure which was originally drafted to illustrate the mtDNA phylogeny; however, as it could not be easily viewed in a journal format, it was not included in the manuscript submitted for publication. Appendix 2 details the AFLP methodology, included since Chapter 2 contains only a brief description of the methods used (as it was submitted for publication), and the AFLP method is highly dependent on consistent technique and quality reagents. Appendix 3 includes the *Mhc* sequence alignments reported in Chapter 3.

[1] Author list: O'Bryan, Dana M.; Zhenyu Xie; Yuxiang Wang; Jizeng Du; Colin J. Brauner; Jeffrey G. Richards; Chris M. Wood; Xeuqun Chen; Brent W. Murray

2 PHYLOGEOGRAPHY AND CONSERVATION GENETICS OF LAKE QINGHAI SCALELESS CARP (*GYMNOCYPRIS PRZEWALSKII*)

2.1 ABSTRACT

Lake Qinghai is a high pH, high elevation, saline (9-11 ppt) lake located on the Qinghai-Tibetan Plateau. The threatened endemic fish, *Gymnocypris przewalskii przewalskii*, has adapted ecologically and physiologically to the extreme environment of the lake. This species is unique among stenohaline cyprinids in that it migrates from the saline lake to adjoining freshwater rivers to spawn. A subspecies, *G. przewalskii ganzihonensis,* resides in the Ganzi River, a geomorphologically distinct tributary that lacks connection to the lake. The objective of this study was to examine the spatial genetic relationships within the Lake Qinghai system, determining whether genetic evidence supports the current taxonomy of *G. p. przewalskii* and *G. p. ganzihonensis* and whether *G. p. przewalskii* are returning to their natal rivers to spawn. Comparison of mitochondrial (control region) variation (42 haplotypes in 203 fish) of *G. przewalskii* with an ancestral species, *Gymnocypris eckloni* (10 haplotypes in 23 fish) indicated no haplotype sharing, but incomplete lineage sorting. Consistent with subspecies status, an AMOVA indicated that the Ganzi River population is significantly different from all other river populations ($F_{ST} = 0.1671$, $p < 0.0001$). No genetic structure was found among the other rivers in the Lake Qinghai watershed. However, an AMOVA of AFLP loci revealed significant genetic differences between most spawning populations ($F_{ST} = 0.0721$, $p < 0.0001$). Both mitochondrial and AFLP data found significant differences among *G. p. przewalskii*, *G. p. ganzihonensis*, and *G. eckloni* (F_{ST} values of 0.1959 and 0.1431, respectively, $p < 0.0001$). Consistent with the incomplete lineage sorting, STRUCTURE analysis of AFLP loci showed evidence of five clusters. One cluster was shared among all sample locations, one was unique to *G. p. ganzihonensis* and *G. eckloni*, and the others were mostly found in *G. p. przewalskii.* Genetic

evidence therefore supported the current taxonomy, including the subspecies status of the Ganzi river fish, and was consistent with natal homing of most Lake Qinghai populations. These findings have significant implications for the conservation and management of this unique and threatened species. We conclude that *G. p. przewalskii* should be treated as a single population for conservation purposes. However, exchangeability of the populations should not be used to promote homogenization of fish spawning in the different rivers. As some degree of genetic divergence was detected in this study, we recommend that the spawning groups be treated as separate management units.

2.2 INTRODUCTION

Lake Qinghai is a saltwater lake (9-11 ppt) located within the Chaidamu Basin of the Qinghai-Tibetan Plateau of western China at an elevation of 3200 m. The lake initially formed during the late Pliocene to mid Pleistocene (about 500,000 years before present) as a result of the downfaulting of the Tibetan Plateau (Bian *et al.*, 2000). Although the ancestral lake originally drained into the Yellow River, the connection was severed approximately 130,000 to 150,000 years ago after an uplift of the plateau (termed the Gonghe Movement). This turned the drainage inward, forming the endorheic basin that comprises the lake today (Zheng, 1997; Li and Fang, 1998).

Although it is the largest lake in China (surface area 4600 km^2), few fish species reside in Lake Qinghai, in part due to extreme conditions (high pH, high saline) (Zhu and Wu, 1975; Zheng, 1997). The most economically important of the fish species is the endemic naked or scaleless carp, *Gymnocypris przewalskii przewalskii* (Kessler) (Cyprinidae: Schizothoracinae). Scaleless carp were fished commercially from the 1950s until the 1990s, when a moratorium was put in place. As the fish are very slow growing, taking approximately 7 years to reach sexual maturity and 10 years to reach a marketable weight of 500 g, stocks have not recovered from overexploitation (Walker *et al.*, 1996). *G. przewalskii* is listed as "Grade B" on a list of species of concern by the China Biodiversity Conservation Action Plan (Zhan, 1994; Yu and Chen, 1998). A small government fishery continues to operate and there have been reports of ongoing poaching.

Scaleless carp are unique among cyprinids in that they undergo an annual migration from the saline lake to adjoining freshwater rivers to spawn, after which, they return to the lake (Walker *et al.*, 1996). The transition from saline water to freshwater and back to saline water

presents significant physiological challenges to the fish, to which they are acclimated in a unique manner. Direct transfer of river fish to lake water caused significant physiological changes while allowing Na^+, Cl^-, and K^+ plasma concentrations to equilibrate with the saline lake water (within 12 hours of transfer from river water). Mg^{2+} plasma levels increased considerably after transfer, though levels were still kept approximately 3 times lower than the lake water, and plasma ammonia levels rose five-fold. It is believed that the overall changes to blood plasma concentrations allow the fish to take a 'metabolic holiday' while residing in the lake as the energy required for ionoregulation and osmoregulation were reduced (Wood *et al.*, 2007).

In addition to the difficulty in recovering from previous overexploitation of stocks, scaleless carp have been presented with two additional challenges. First, freshwater tributaries have been dammed or diverted for agricultural purposes. In combination with decreased rainfall, this has resulted in the lake becoming progressively shallower. Water levels have fallen at a rate of approximately 10-12 cm per year over the last 50 years, resulting in increasing salinity levels (Walker and Yang, 1999). Salinity increases have occurred since 1978, and current data indicates that this increase is accelerating (Wang *et al.*, 2003; Yang *et al.*, 2005; Wood *et al.*, 2007). Although scaleless carp are currently able to physiologically tolerate the salinity of the lake, experiments have shown that a slight increase in salinity levels (4 ppt) would be toxic (Brauner *et al.*, 2008). Second, the damming and diversion of rivers has resulted in a significant loss of spawning habitat. Irrigation weirs were constructed on the Shaliu and Haerga Rivers in the 1950s, restricting access to spawning locations upstream (Walker and Yang, 1999). The Quanji River was later dammed approximately three kilometres from the mouth of the lake, and a fish ladder was not installed until 2006.

Based on morphological characters, a subspecies of the scaleless carp, *Gymnocypris przewalskii ganzihonenesis*, has been identified in the Ganzi River (Zhu and Wu, 1975). While ancient Chinese historical records indicate that the Ganzi River may have once flowed directly into Lake Qinghai, a 1964 survey found that it lacked direct connection to the lake, likely due the shrinking of the lake shoreline and low flows in the lower reaches of the river (Zhu and Wu, 1975). The current situation, which is heavily impacted by irrigation, is similar to the conditions in 1964. The lower and middle reaches consist of marshes with very low flows, and large dunes block direct surface connection to the lake. As the current status of the population is unknown, an important goal of the present study was to determine the genetic relationship between *G. p. przewalskii*, *G. p. ganzihonensis* and the closely related Yellow River species, *Gymnocypris eckloni* Herzenstein. The Ganzi River has been heavily impacted by anthropogenic disturbances (*e.g.* diversion of water for agriculture) and information on this isolated fish population is critical to initiation of conservation actions.

Several different genetic markers have been employed in the study of fish conservation. This study used the control region of mitochondrial DNA in combination with Amplified Fragment Length Polymorphisms (AFLPs). There are an increasing number of fish studies that combine mitochondrial variation with nuclear markers, such as AFLPs (Young *et al.*, 2001; Seehausen *et al.*, 2003; Ishikawa *et al.*, 2004; Mock and Miller, 2005). The mtDNA control region is generally selectively neutral (non-coding) and provides information on non-selective processes, such as mutation and drift. Mitochondrial DNA is frequently used in population genetic structure studies (Brunner *et al.*, 1998; Stepien and Faber, 1998; Taylor *et al.*, 2001; Gharrett *et al.*, 2005) and to assess subspecies (Tabata and Taniguchi, 2000; April and Turgeon, 2006). AFLPs, developed by Vos and colleagues (1995), allow analysis of multiple loci across

the entire nuclear genome. These markers are capable of identifying relationships at the species or population level, providing the resolving power to detect the early divergence of lineages (Sullivan *et al.*, 2004). AFLPs are, therefore, especially well suited to studies of organisms with very little prior genetic information, as in *G. przewalskii.*

The main objective of this study was to assess the spatial genetic structure of *G. przewalskii* within the Lake Qinghai system. Samples were collected from each of the major rivers that drain into Lake Qinghai and from the Ganzi River. Samples of *G. eckloni*, the most closely related species, were collected from the Yellow River. This sampling design allowed genetic evidence for the existence of a Ganzi river subspecies to be placed into the context of genetic divergence from the closest outgroup to the Lake Qinghai fish. It also assessed the degree of structure found among fish caught in the rivers draining into the lake. Identification of stock structure within the lake would indicate that the scaleless carp preferentially spawn in the same rivers in which they were born (natal homing). The ability to identify different stocks within Lake Qinghai is also extremely important for the conservation and management of scaleless carp populations both within the lake and in each river. Some tributaries, and associated spawning habitat, may be lost due to falling lake levels, agricultural divergence, and damming. Findings from this study will have significant implications for the conservation and management of this unique and threatened species.

2.3 METHODS

2.3.1 Field Sampling

Lake Qinghai was visited in June, 2002, 2004, and 2005 during the scaleless carp spawning run (Figure 2.1). Fish samples were collected under permits issued by local and

national authorities and in accordance with national animal care regulations. An animal care and use approval was obtained of this research from Queen's University, Ontario, Canada. Adult fish were collected from most rivers using a throw net; however, juvenile fish were collected from the Ganzi River as adult fish were not observed. Caudal fin clips [preserved in a DMSO salt solution (Seutin *et al.*, 1991)] were mostly taken from live fish, which were then released, although a small number of fish were lethally sampled. Samples were taken from one location each in the Heima, Quanji, Shaliu, Haerga, and Ganzi Rivers, and from two locations in the Buha River (the largest tributary). Sample numbers and map coordinates are shown in Figure 2.1. *G. eckloni* were collected by a local fisherman from the upper reaches of the Yellow River near the town of Madoi in June 2004 (n = 23 and 38 for mtDNA and AFLP analyses respectively).

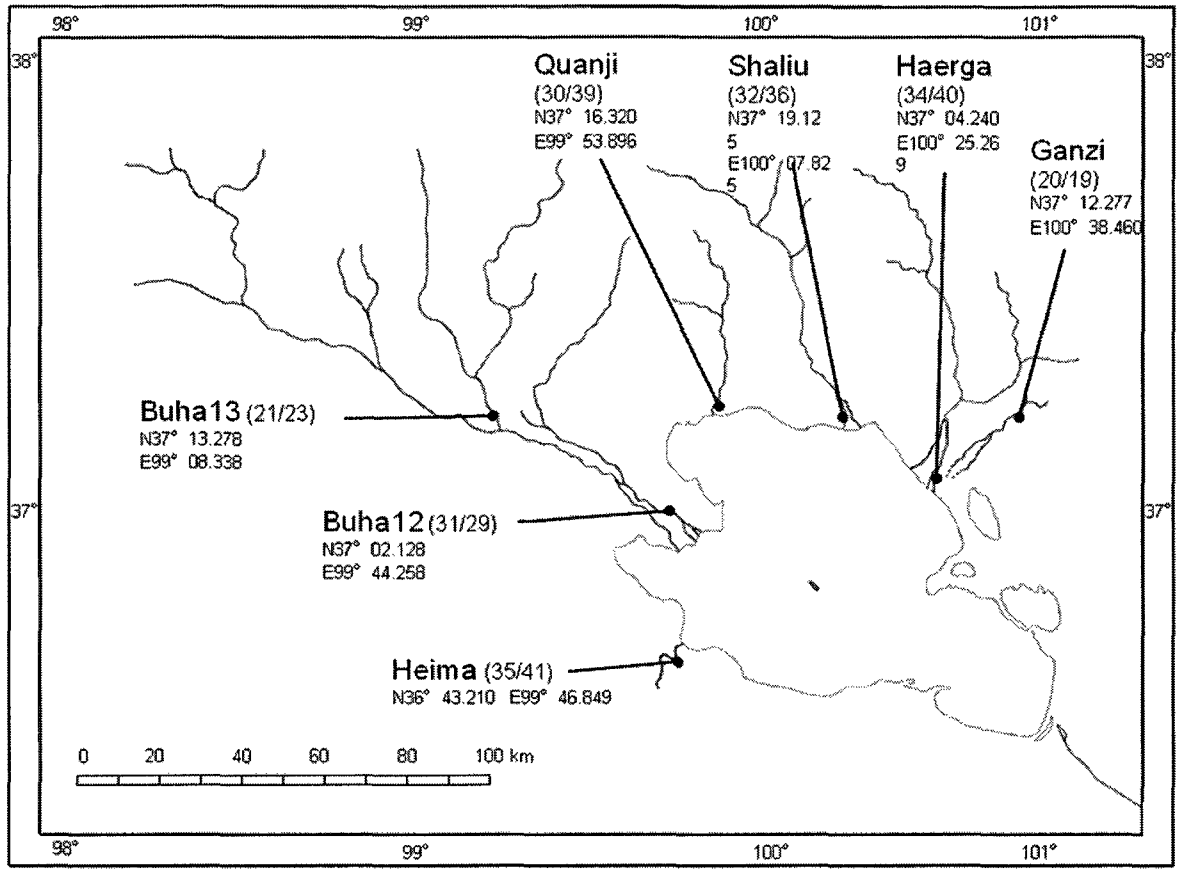


Figure 2.1: Map of the Lake Qinghai watershed showing sampling locations of *G. p. przewalskii* and *G. p. ganzihonesis*. For each location the latitude and longitude and the sample size used for the mtDNA and AFLP analyses, respectively, are shown.

2.3.2 Laboratory Work

DNA was extracted from tissue samples using a standard phenol/chloroform method followed by ethanol precipitation and resuspension in a standard Tris-EDTA solution (Sambrook and Russell, 2001).

2.3.2.1 mtDNA Analysis

Approximately 50 ng of genomic DNA was added to 24 μL of PCR master mix containing 1X PCR buffer (Invitrogen, Carlsbad, CA), 250 μM of each dNTP, 2mM $MgCl_2$, 0.2 μM of each primer (ProL1 and CRH1; Table 2.1), and 1 unit of *Taq* (Invitrogen). The

amplification program was 94°C for 4 minutes, 47°C for 1 minute, and 72°C for 2 minutes, followed by 39 cycles of 94°C for 30 seconds, 47°C for 30 seconds, and 72°C for 90 seconds, and ending with 72° for 8 minutes. Amplifications were carried out on DYAD or PTC-100 thermal cyclers (MJ Research, Inc., Waltham, MA). The resulting DNA was purified using the Wizard® PCR Preps DNA Purification System (Promega, Madison, WI). DNA sequencing was performed using the DTCS Quick Start Kit (Beckman Coulter, Inc., Fullerton, CA) using 1.6 μM SProl or SH1 primers (Table 2.1). Sequence fragments were separated on the CEQ 8000 Genetic Analysis System (Beckman Coulter) with method LFR-1.

Table 2.1: Oligonucleotides used in mitochondrial and AFLP analyses.

mtDNA Oligonucleotides	
ProL1	5' CCTAGCTCCCAAAGCCAGA 3'
CRH1	5' TGTTGGCTGATACGTTCTTGG 3'
SProL1	5' TTCTGAACTAAACTATTCTCTGG 3'
SH1	5' CCATTTACTGAGCGTAGGG 3'
AFLP Oligonucleotides	
Adapters	
*Eco*RI-adapter-fwd	5' CTCGTAGACTGCGTACC 3'
*Eco*RI-adapter-rev	5' AATTGGTACGCAGTCTAC 3'
*Mse*I-adapter-fwd	5' GACGATGAGTCCTGAG 3'
*Mse*I-adapter-rev	5' TACTCAGGACTCAT 3'
Pre-selective Primers	
*Eco*RI-A	5' GACTGCGTACCAATTCA 3'
*Mse*I-C	5' GATGAGTCCTGAGTAAC 3'
Selective Primers	
E-AAG	5' GACTGCGTACCAATTCAAG 3'
E-ACG	5' GACTGCGTACCAATTCACG 3'
M-CAA	5' GATGAGTCCTGAGTAACAA 3'
M-CAC	5' GATGAGTCCTGAGTAACAC 3'
M-CAG	5' GATGAGTCCTGAGTAACAG 3'

2.3.2.2 mtDNA Data Analysis

Sequence chromatograms were viewed and edited in Sequencher 4.2 (Gene Codes, Ann Arbor, MI). To confirm variation, haplotypes were sequenced in both directions (with both SProL1 and SH1 primers) and, in many cases, multiple individuals (see Figure 2.2). Sequences

reported here have been deposited in the Genbank database under accession numbers EU150135-EU150186.

Sequences were aligned with ClustalW implemented in MEGA 3.1 (Kumar *et al.*, 2004). An Analysis of Molecular Variance (AMOVA) was performed in Arlequin 3.11 (DNA data type; genotypic data set to 0; 1000 permutations) (Excoffier *et al.*, 2005). Three hierarchical analyses were conducted by first grouping the data in accordance with the currently accepted taxonomy and then sequentially removing the *G. eckloni* and *G. p. ganzihonensis* groups. Population statistics, including haplotype diversity, nucleotide diversity, and estimates of θ ($\theta = 2N_f\mu$, where N_f = female effective population size and μ = mutation rate) were also calculated in Arlequin. Phylogenetic relationships among the haplotypes were studied using neighbor-joining (Kumar *et al.*, 2004), parsimony networks (Clement *et al.*, 2000) and maximum likelihood methods (Guindon and Gascuel, 2003). jModeltest 0.1.1 (Posada, 2008) was used to find the best-fit model of nucleotide substitution and to calculate a maximum likelihood model-averaged phylogeny.

2.3.2.3 Amplified Fragment Length Polymorphism (AFLP) Analysis

The AFLP protocol was based on Vos *et al.* (1995) and Hawkins *et al.* (2005) with modifications (Hayashi *et al.*, 2005). A list of locations and numbers of samples used is included in Figure 2.1. All adapter and primer sequences are listed in Table 2.1

Two hundred nanograms of high quality genomic DNA was digested (3 hours at 37°C) with 10 units each of *Mse*I and *Eco*RI, 1X *Eco*RI Restriction Buffer, and 2 µg of Bovine Serum Albumin (BSA) (New England Biolabs, Ipswich, MA) in a 20 µL reaction. A 20 µL ligation master mix containing 1X T4 Ligase Buffer (New England Biolabs), 3 units (NE Units) of T4

DNA Ligase, and 1.5 μL each of 50 μM adapters (*Mse*I and *EcoR*I) was added to the restriction digest and incubated at 16°C for at least 16 hours.

Preselective amplifications were performed by combining 5 μL of diluted digestion-ligation product (1:5 in nuclease-free water) with 20 μL of a preselective master mix containing 1X PCR Buffer, 1.5mM $MgCl_2$, 200 μM of each dNTP, 0.8μM each of the *EcoR*I-A and *Mse*I-C preselective primers, and 1.25 units of *Taq* polymerase (Invitrogen). The amplification program was 72°C for 2 minutes, followed by 20 cycles of 94°C for 30 seconds, 56°C for 1 minute, and 72°C for 1 minute, and ending with 60°C for 30 minutes.

Selective amplification was subsequently performed by combining 5 μL of diluted preamplification product (1:10 in nuclease-free water) with 15 μL of a selective master mix containing 1X PCR Buffer, 1.5mM $MgCl_2$, 200 μM of each dNTP, 0.4 μM each of the *EcoR*I and *Mse*I selective primers, and 1 unit of Platinum *Taq* (Invitrogen). Amplifications were carried out with an initial activation of 94°C for 30 seconds, followed by 12 cycles of 94°C for 30 seconds, 65°C (decreased by 0.7°C per cycle) for 30 seconds, and 72°C for 1 minute. The preceding cycles were followed by 32 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute, ending with a final extension of 72°C for 10 minutes. The selective primer combinations used in this study were E-ACG/M-CAA, E-ACG/M-CAG, and E-AAG/M-CAC.

Repeatability was checked by randomly choosing samples to run in triplicate prior to the digestion-ligation procedure (one sample in forty was run in triplicate). Repeatability was further assured by running each digestion-ligation product in duplicate through the pre-amplification and selective amplification procedures.

2.3.2.4 AFLP Data Analysis

AFLP fragments were sized using the CEQ8000 Genetic Analysis system (Beckman Coulter), and viewed in the Fragment Analysis Module (600 bp size standard; quartic model; PA ver.1 dye mobility calibration). Only reliable loci, which had strong intensities, were clear and crisp, and greater than 80 base pairs in size were included in the analysis. Any fragments which were observed in blank samples were removed from analysis, although their presence does not necessarily indicate contamination, but likely results from primer-dimers and interactions between adapters (Bensch and Åkesson, 2005). To further increase confidence in the data set, each sample's replicate was scored, and if a fragment did not occur in both replicates, it was also removed from the analysis, although this rarely occurred (fewer than 5% of runs contained a fragment that was removed). Fragments were binned using the AFLP analysis feature within the Fragment Analysis module (1 nucleotide bin width[2] and the Y threshold set to 0). Similar fragment sizes were assumed to represent homology. The data were represented as a binary matrix (1 indicates the presence of the fragment while 0 indicates its absence).

AFLP data were examined on an individual level using STRUCTURE 2.2 adapted to dominant markers (Falush *et al.*, 2007). Correlated allele frequencies and admixed population origins were assumed. Burn-in and run lengths were set to 10000. The number of inferred groups (k) ranged from 2 to 8, and each possible k value was run 20 times. The "best" value of k was objectively determined using the method of Evanno *et al.* (2005). Hierarchical AMOVA and pairwise F_{ST} comparisons were carried out using Arlequin 3.11 (RFLP data type; genotypic data and gametic phase set to 0; recessive data set to 1; 1000 permutations; assumed group structure based on current taxonomy) (Excoffier *et al.*, 2005). Principle Coordinate Analysis of

[2] Short width was set in order to maintain complete control over the binning of fragments

pairwise sample location genetic distance (Nei, 1972) was carried out in GenAlEx 6 (Peakall and Smouse, 2006). Percent polymorphism was calculated in Tools for Population Genetic Analysis 1.3 (Miller, 1997).

As adult fish were not observed in the Ganzi River at the time of sampling, only juvenile fish were collected. To test for the possible effects of sampling siblings (and thereby skewing the results), the average pairwise genetic distance (Nei, 1972) for AFLP data was compared for individuals with shared mtDNA haplotypes (which are more likely to be siblings) versus those that do not share mtDNA haplotypes. If siblings were sampled (and therefore the entire Ganzi river dataset would be biased), the expectation is that the average genetic distance will be less for individuals that share haplotypes compared to those that do not.

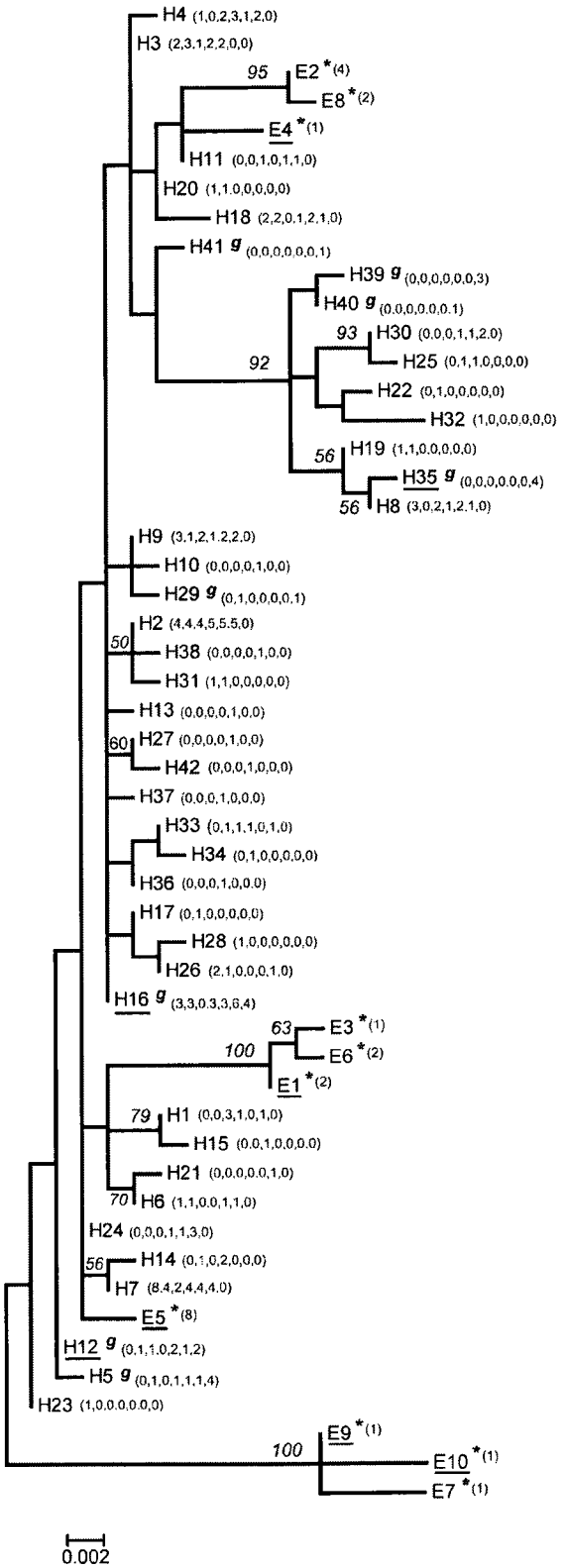
2.4 RESULTS

2.4.1 mtDNA Analysis

Sequences reported in a preliminary analysis [Genbank accession numbers AY850398 and AY850397 (Xie *et al.*, 2006)] were combined with sequences from this study, resulting in the identification of 42 distinct scaleless carp haplotypes (Figure 2.2). Ten *G. eckloni* haplotypes were also identified, none of which were shared with *G. przewalskii*. From a 701 bp alignment of 52 haplotypes, jMODELTEST was used to find the best-fit model of nucleotide evolution (Posada, 2008). The best-fit model was used to estimate a maximum likelihood tree (Figure 2.2) and a rank weighting of models to estimate a model-averaged maximum likelihood tree. The model-averaged method (Posada, 2008) investigates the sensitivity of the tree topology to the model used. Our results indicate that the tree is relatively insensitive to the model used. All nodes were supported by model-averaged weights of 1.0, indicating the same topology was

found in all the top weighted best-fit models.

Figure 2.2: Maximum likelihood tree of G. przewalskii and G. eckloni mitochondrial DNA haplotypes identified by sequencing the control region. The best fit model of nucleotide substitutions identified using the Bayesian Information Criterion (BIC) implemented in jModeltest 0.1.1 (Posada, 2008), HKY with unequal base frequencies and a proportion of invariable sites, was used by program PHYML (Guindon and Gascuel, 2003) to estimate the above phylogeny and calculate bootstrap values. Bootstrap values greater than 50% are shown. A model-averaged phylogeny (Posada, 2008) using relative model weights based on either the BIC or Akaike Information Criterion is identical in topology to the tree shown with all nodes having a model support of 1.0. 'H' indicates *G. przewalskii* and 'E' (with asterisks) indicates *G. eckloni* haplotypes. Haplotypes found in the Ganzi River are indicated by 'g' superscript. The sample size of each haplotype is shown in brackets. For the *G. przewalskii* haplotypes the sample size in each sample location is shown in order of Heima, Buha12, Buha13, Quanji, Shaliu, Haerga and Ganzi, respectively. Underlined haplotypes are found in samples that are members of an AFLP cluster found in the Yellow and Ganzi rivers (see Figure 2.3).



The maximum likelihood tree of the *G. przewalskii* and *G. eckloni* haplotypes showed incomplete lineage sorting (*i.e.*, the *G. eckloni* haplotypes do not form a monophyletic group). Incomplete lineage sorting was also noted when a parsimony network (Figure A1.1) and neighbor joining trees (not shown) were constructed. The topology of the maximum likelihood tree was very similar to the parsimony network in that many haplotypes differed by a single nucleotide change and polytomies were common. Basal haplotypes (inferred ancestral haplotypes from which the others branches radiate) are shown at nodes with no branch length. There are three major clusters and one singleton of *G. eckloni* haplotypes that occur among clusters of *G. przewalskii* haplotypes. There do not appear to be any clear geographical relationships among the *G. przewalskii* haplotypes. Of the 42 scaleless carp haplotypes, 19 were unique to only one population (Table 2.2). Of the 8 haplotypes detected in the Ganzi River fish, four (50%) were unique to that river, even though the sample size was smaller than from the other rivers that had between 22% to 6% unique haplotypes. The four Ganzi haplotypes that were shared with the other populations generally represented frequently occurring haplotypes in the lake (Figure 2.2).

Table 2.2: Number of mtDNA haplotypes and number of unique haplotypes (percent in brackets) found within G. przewalskii sampling locations.

	Heima	Buha 12	Buha 13	Quanji	Shaliu	Haerga	Ganzi
Haplotype number	16	20	12	17	18	17	8
Unique haplotypes	3 (19)	3 (15)	1 (8)	3 (18)	4 (22)	1 (6)	4 (50)

AMOVA results indicated significant differences among *G. p. przewalskii*, *G. p. ganzihonensis*, and *G. eckloni* ($F_{ST} = 0.1959$, $F_{CT} = 0.2083$) when populations were grouped according to the current taxonomy (Table 2.3). Significant differences ($p < .0001$) were also

noted when comparing *G. p. ganzihonensis* to the lake fish (results not shown: $F_{ST} = 0.1671$, $F_{CT} = 0.1781$). Removal of *G. p. ganzihonesis* to examine only differences among sample locations within the current lake drainage revealed no evidence for genetic structure (*i.e.*, F_{ST} not significantly different from 0). These relationships are further illustrated by pairwise comparisons of F_{ST} (among sample locations) values where *G. p. ganzihonensis* and *G. eckloni* were found to be significantly different from all *G. p. przewalskii* samples (pairwise F_{ST} from 0.1003 to 0.1923) (Table 2.4). Pairwise comparisons between the *G. p. przewalskii* samples from the other rivers, however, did not exhibit significant differences (*i.e.*, pairwise F_{ST} values ranged from -0.02436 to 0.0063; all p values > 0.25).

Table 2.3: Results of hierarchical AMOVA for both mtDNA and AFLP markers with populations (n=8) grouped according to the currently accepted taxonomy (3 groups: *G. p. przewalskii* (6), *G. p. ganzihonensis* (1), *G. eckloni* (1)). F statistics were generated in Arlequin 3.11 (Excoffier et al. 2005). Degrees of freedom (df), Sum of Squares (SS), Variance components (Var), Percent variance (% Var) and the p-values (p) of % Var are shown. F_{CT} is the variance among groups. F_{ST} with group structure is the fraction of total variance found among populations within groups plus the variance among groups. $F_{ST} = 0.1959$, $p < .0001$ for mtDNA and 0.1431, $p < .0001$ for AFLP markers.

	mtDNA					**AFLP**				
	df	SS	Var	% Var	p*	Df	SS	Var	% Var	p*
Among Groups	2	51.00	0.648	20.84	.0000 (.0420)	2	79.62	0.480	7.76	.0273 (.0420)
Among Pop within Groups	5	6.62	0.039	-1.25	.9658 (.0000)	5	96.23	0.405	6.55	.0000 (.0000)
Within Pop	218	544.83	2.499	80.41	.0000 (.0000)	257	1361.8	5.217	85.69	.0000 (.0000)

*Given the specified group structure, there is a high probability of recreating the tested structure; therefore, p values where the random values are greater than the observed values are reported, with p values of random values equal to the observed in brackets.

Table 2.4: Pairwise F_{ST} values for *G. przewalskii* and *G. eckloni* sampling locations from mtDNA haplotypes (above the diagonal) and AFLP genotypes (below the diagonal). Only significant values are shown (all $p < 0.0001$, except those marked with an asterisk where $p < 0.05$) and non-significant values are indicated with a "*NS*".

	Heima	Buha 12	Buha 13	Quanji	Shaliu	Haerga	Ganzi	Yellow
Heima	-	*NS*	*NS*	*NS*	*NS*	*NS*	0.1251*	0.1640
Buha 12	0.1532	-	*NS*	*NS*	*NS*	*NS*	0.1565	0.1696
Buha 13	0.1165	0.1356	-	*NS*	*NS*	*NS*	0.1003*	0.1302
Quanji	0.0367	0.0970	0.0435	-	*NS*	*NS*	0.1923	0.1874
Shaliu	0.0303	0.1504	0.0547	*NS*	-	*NS*	0.1524	0.1756
Haerga	0.0656	0.1341	0.0588	0.0157*	0.0284	-	0.1691	0.1771
Ganzi	0.1654	0.1571	0.2484	0.1343	0.1884	0.1564	-	0.1801
Yellow	0.1150	0.1674	0.1940	0.1089	0.1199	0.1025	0.0956	-

Population statistics were calculated with samples grouped based on the current taxonomy (Table 2.5). θ values ranged from 3.94 to 14.30 and gave estimates of female effective population sizes ranging from 19,700 to 71,500.

Table 2.5: Population statistics generated using mtDNA (h = haplotype diversity, π = nucleotide diversity, $\theta_{(k)}$ = Theta estimated from the expected number of alleles (k) and sample size (n) (Ewen, 1972), $\theta_{(Hom)}$ = Theta estimated based on the expected homozygosity (Hom) (Zouros, 1979; Chakroborty and Weiss, 1991), $\theta_{(S)}$ = Theta estimated from the number of segregating sites (s) and sample size (n) (Watterson, 1975), $\theta_{(\pi)}$ = Theta estimated from the mean number of pairwise differences (Tajima, 1983), N_f = range of female effective population sizes calculated using previous estimates of θ).

Species	h ± SD	π ± SD	$\theta_{(k)}$ 95% CI	$\theta_{(Hom)}$ ± SD	$\theta_{(S)}$ ± SD	$\theta_{(\pi)}$ ± SD	N_f
G. p. przewalskii	0.9340 ± 0.0081	0.0060 ± 0.0031	14.30, 9.73-20.69	12.65 ± 1.80	6.22 ± 1.68	4.21 ± 2.32	71,500-21,050
G. p. ganzihonensis	0.8842 ± 0.0361	0.0084 ± 0.0046	4.44, 1.84-10.38	6.43 ± 2.50	3.94 ± 1.65	5.86 ± 3.26	32,150–19,700
G. eckloni	0.8538 ± 0.0571	0.0147 ± 0.0077	6.18, 2.75-13.62	4.79 ± 2.40	10.57 ± 3.78	10.26 ± 5.42	52,850–23,950

2.4.2 AFLP Analysis

A total of 41 polymorphic loci were included in this analysis, with an average polymorphism across both species of 77.3%. All 41 loci were polymorphic in the *G. eckloni* samples while 39 were polymorphic within the *G przewalskii* samples. The Ganzi River sample location exhibited the smallest level of polymorphism (Table 2.6).

Table 2.6: Percent polymorphism of AFLP fragments in *Gymnocypris* samples from each sampling location.

Sampling Location	% Polymorphic Loci
Heima	66.0
Buha 12	62.3
Buha 13	58.5
Quanji	64.2
Shaliu	62.3
Haerga	71.7
Ganzi	56.6
Yellow	67.9

STRUCTURE analysis indicated that the greatest rate of change in likelihood occurs between k=3 and k=4. Applying the method of Evanno and colleagues (2000), an optimal k value of 5 was found. The results of the STRUCTURE analysis are best illustrated by comparing the results of k=3 and k=5 (Figure 2.3). No direct association of sampling locations with inferred groups was observed. *G. p. ganzihonensis* and *G. eckloni* individuals were primarily contained within two shared clusters, one mostly unique to these locations (green), and one shared with all *G. p. przewalskii* sample locations (blue). In contrast, the six *G. p. przewalskii* sample locations each had a number of individuals belonging to a third cluster at k=3 (red). Increasing k values to 4 or 5 split this third cluster into additional groups (red, yellow and pink) while primarily maintaining the integrity of the first two. At k=5, two *G. eckloni* samples were included into the red and pink clusters commonly found in the *G. p. przewalskii* sample locations. Stock structure among *G. p. przewalskii* sample locations may also be illustrated by the frequency differences of cluster membership at each location. For example the percentage of the yellow cluster, unique to *G. p. przewalskii* samples, varies from 2-43% among these locations, with the greatest percentage being found in the two Buha river sites (24 and 43%).

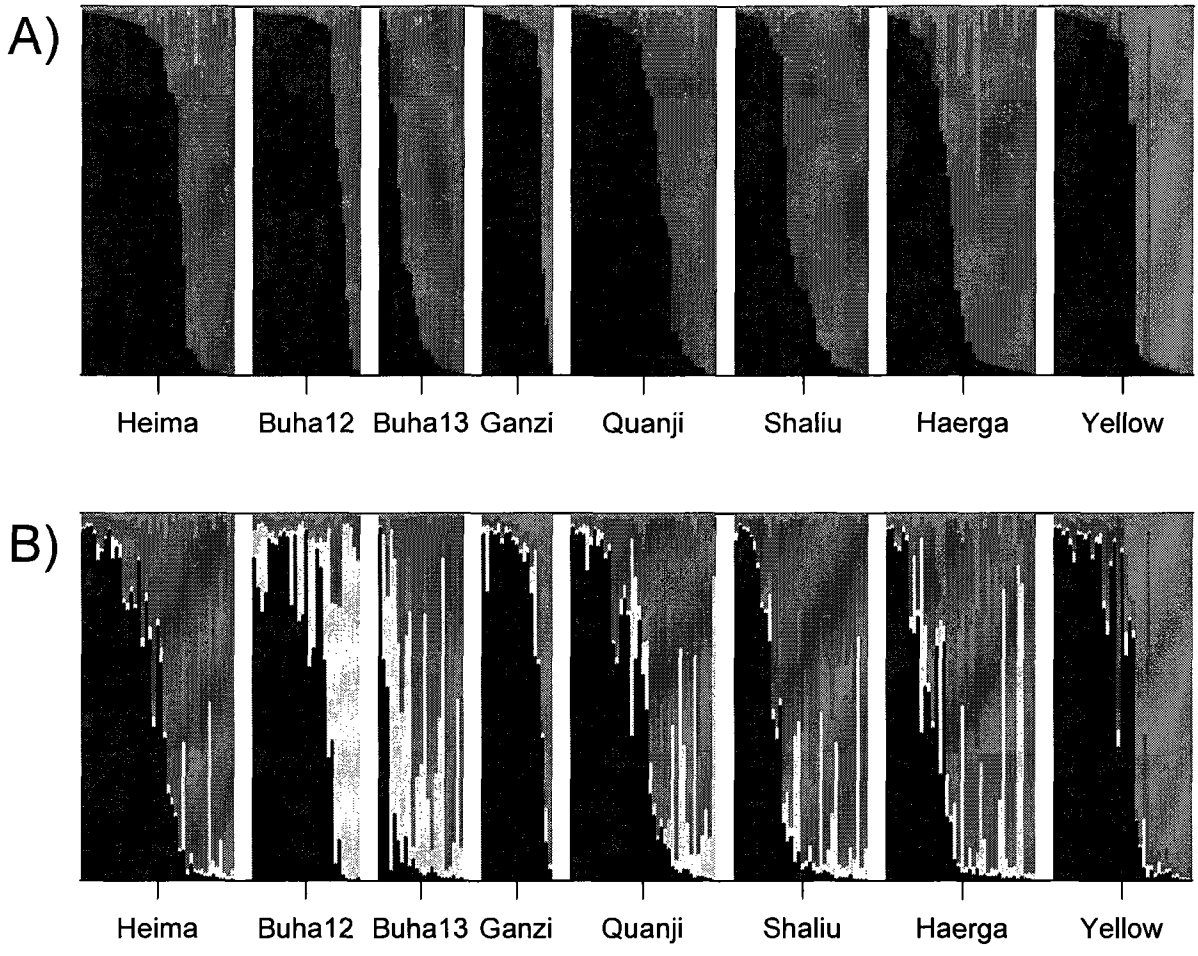


Figure 2.3: Genetic structure estimated from AFLP markers for k = 3 (A) and k=5 (B) using STRUCTURE 2.2 (Falush *et al.*, 2007). *Gymnocypris* individuals are grouped by sampling location; Heima, Buha 12, Buha 13, Quanji, Shaliu and Haerga for *G. p. przewalskii*; Ganzi for *G. p. ganzihonesis*; and Yellow for *G. eckloni.* Sample order is the same at both k=3 and k=5. Cluster membership was determined using the largest coloured bar for each individual.

One cluster (green) was unique to the *G. eckloni* and *G. p. ganzihonesis* samples at k=5. The members of this cluster were compared to the results of the mtDNA analysis. Underlined haplotypes (Figure 2.2) were found in individuals with AFLP genotypes that belong to this cluster. No correlation between mtDNA and AFLP genotypes was noted. In the three *G. p. ganzihonesis* samples belonging to this cluster, two had common, but divergent, *G p. przewalskii*

mtDNA haplotypes while one was unique to the Ganzi. Similarly, *G eckloni* mtDNA haplotypes in this cluster were found in each of the four *G. eckloni* mtDNA lineages.

Similar to the mtDNA results, AMOVA of the AFLP data showed significant differences ($p < 0.05$) among all groups (Table 2.3: $F_{ST} = 0.1431$, $F_{CT} = 0.0776$) and between *G. p. przewalskii* and *G. p. ganzihonensis* (Results not shown: $F_{ST} = 0.1681$, $F_{CT} = 0.1031$) when grouped according to the current taxonomy. In contrast, significant differences among sample locations within *G. p. przewalskii* were also found ($F_{ST} = 0.0721$, $p < .0001$). Pairwise comparisons of F_{ST} values between all sampling locations found all but the Quanji and Shaliu samples to be significantly different from each other (Table 2.4). Pairwise F_{ST} values for the Ganzi River fish were generally higher (0.1343 to 0.2484) when compared to the values found between *G. p. przewalskii* sampling locations.

To further explore the differences among sample locations, a Principle Coordinates Analysis of pairwise genetic distances was conducted (Figure 2.4). Most of the Lake Qinghai locations were found to group in the same quadrant, while the Buha sample locations were more distant and in a separate quadrant. The *G. p. ganzihonensis,* Ganzi, and *G. eckloni*, Yellow, sample locations were also each in separate quadrants and distant from all other locations.

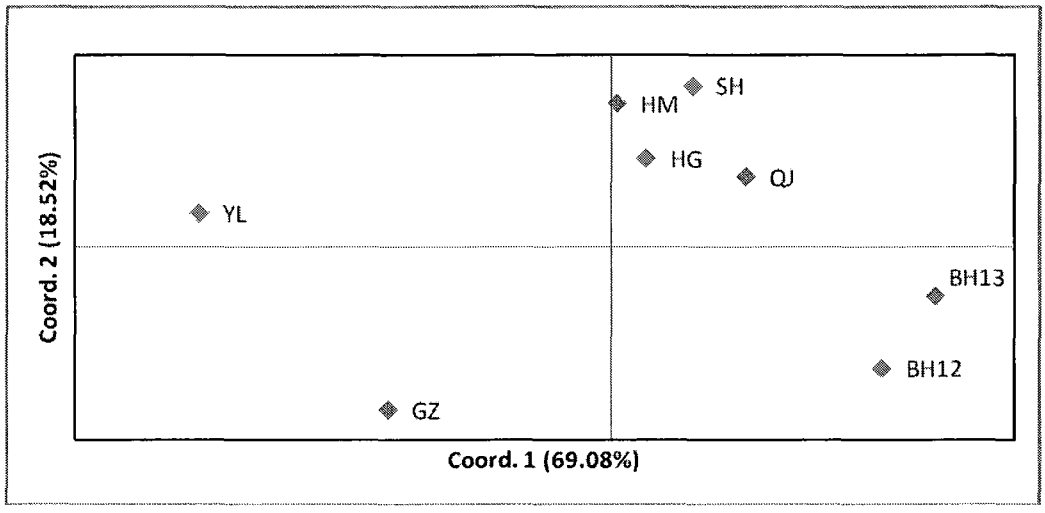


Figure 2.4: Principle Coordinates Analysis of pairwise genetic distance matrix (Nei's Genetic Distance) among sampling locations using GenAlEx 6 (Peakall and Smouse, 2006). The first two coordinates explain 87.61% of the variation. *Gymnocypris* sampling locationa are Heima, Buha 12, Buha 13, Quanji, Shaliu and Haerga for *G. p. przewalskii*; Ganzi for *G. p. ganzihonesis*; and Yellow for *G. eckloni.*

2.4.3 Test for the effects of sampling juveniles in the Ganzi River

When sampling small populations of juveniles, collecting sibling groups is a possibility. If siblings were sampled in this case, it is expected that the average genetic distance (calculated using AFLP data, not shown) would be less for individuals that shared mtDNA haplotypes compared to those that did not. However, the average pairwise genetic distance (calculated using AFLP data) for individuals that shared mtDNA haplotypes (0.011) was not significantly different from the average genetic distance for individuals that did not share mtDNA haplotypes (0.012) ($t = -1.131$, $p = 0.301$), indicating that sibling groups were not likely sampled in the Ganzi River.

2.5 DISCUSSION

The results of both the mtDNA and AFLP analysis indicated a population structure that is consistent with the current taxonomy. Among sample locations collected within the existing Lake Qinghai watershed, only the AFLP data indicated the existence of genetic structure. Crandall and colleagues (2000) stressed that both genetic and ecological data (both recent and

historic in nature) should be used in defining Evolutionary Significant Units (ESUs). ESUs should lack both ecological and genetic exchangeability. Lack of ecological exchangeability refers to different selective pressures and adaptations between populations. These might include differences in life history traits, habitats, predators and/or morphology. Lack of genetic exchangeability refers to lack of recent gene flow between populations and the observation of genetic differences consistent with geographic divisions.

2.5.1 Relationship between *G. eckloni* and *G. przewalskii*

G. przewalskii and *G. eckloni* do not share any mtDNA control region haplotypes; however, there is incomplete lineage sorting of the haplotypes (Figure 2.2), a result that was also found in previous comparisons of the two species using mitochondrial sequences (Zhao *et al.*, 2009; Zhao *et al.*, 2007; Zhao *et al.*, 2005). Using the high and low values for the *G. p. przewalskii* female effective population size (Table 2.6) and a generation time of 7 years (Walker *et al.*, 1996), the estimated time for all haplotypes to coalesce was 41, 097 to 139, 595 generations, or 287, 686 to 977, 166 years[3]. The estimated molecular divergence time reported using cytochrome b sequences between the fish of Lake Qinghai and the Yellow River also coincided with the estimated time of separation of the two water bodies during the Gonghe movement (Qi *et al.*, 2007). Given that the lowest estimate for all haplotypes in Lake Qinghai or the Yellow River to coalesce was approximately 287,686 years and that the minimum estimated time for separation of the two water bodies is 130,000 years, incomplete lineage sorting in Lake Qinghai is not unexpected. Indeed, this would be consistent with a relatively recent speciation

[3] Using the formula $E(T) = 2Nf\frac{(n-1)}{n}G$, where *E(T)* = estimated time for the coalescence of all haplotypes, N_f = female effective population size, *n* = number of haplotypes, and *G* = generation time. Adapted from Hedrick (2005).

event occurring sometime during the separation of the Yellow River and Lake Qinghai (Zhao *et al.*, 2007; Zhao *et al.*, 2005; Qi *et al.*, 2006).

The results of the STRUCTURE analysis of AFLP genotypes were also consistent with a recent ancestry of the species. As demographic exchange is not possible, the AMOVA results also show a divergence of the nuclear genotypes of the species. Lineage sorting, however, is not complete as noted by the high rate of sharing of polymorphic AFLP markers (39 of 41). Given the larger effective population size for nuclear genes versus mtDNA, lineage sorting of polymorphic AFLP markers is not expected. The sharing of clusters and markers is more than likely a result of a recent common ancestry. Further, the *G. eckloni* samples were caught in the main stream of the Yellow river. The existence of multiple spawning runs within this (and possibly other – see below) sampling location would also further increase the complexity of the structure results.

According to the definition of Moritz (1994), the lack of reciprocal monophyly at mtDNA markers would prevent the distinction of *G. eckloni* and *G. przewalskii* as separate ESUs (or species). This precludes the recognition of ESUs until completion of lineage sorting. Alternatively, Crandall and colleagues (2000) advocate a conservation management strategy that examines genetic and ecological data across two timescales – recent and historic. The examination of genetic or ecological exchangeability on a historic timescale can be problematic; however, the use of two different molecular markers may allow the partitioning of recent and historical phenomena. In this study, mitochondrial DNA may be more indicative of historical relationships, whereas AFLPs, due to the large number of loci examined, may be able to detect more recent population divergence (Egger *et al.*, 2007). *G. eckloni* and *G. przewalskii* were found to be significantly different at both mtDNA and AFLP markers. The two species are also

morphologically distinct and occupy different habitats (*i.e.*, *G. eckloni* permanently reside in the freshwater Yellow River drainage), an indicator of recent ecological non-exchangeability. Therefore, using the principles of genetic and ecological exchangeability (Crandall *et al.*, 2000), *G. eckloni* and *G. przewalskii* should be considered different ESUs, consistent with the current taxonomy.

2.5.2 Ganzi River Subspecies

Both the control region and AFLP markers indicated that the Ganzi River fish are significantly different from all *G. p. przewalskii* populations, consistent with the current subspecies status. AMOVAs of both the control region and AFLP data (Table 2.3 and 2.4) found the Ganzi fish to be significantly different from all sampling locations of *G. p. przewalskii*. As the average genetic distance (calculated with AFLP data) of individuals that shared mtDNA haplotypes was not significantly different from those that did not share haplotypes, it is unlikely that the results in this study were due to the sampling of sibling groups. Consistent with the results of the species level comparison and previous findings (Zhao *et al.*, 2005), the Ganzi population was not found to be monophyletic when the control region was analyzed. However, fewer haplotypes were shared with *G. p. przewalskii* than the degree of haplotype sharing found among *G. p. przewalskii* sampling locations (Table 2.5). Additionally, STRUCTURE analysis revealed the presence of AFLP genotypes within the Ganzi that were more similar to Yellow River fish than to the *G. p. przewalskii* (Figure 2.3). The AFLP data were consistent with the expectations of ongoing lineage sorting at nuclear genes. Further, there was no evidence to support recent genetic exchange, either natural or anthropogenic, with the Yellow river species. No Yellow river mtDNA haplotypes were found in the Ganzi. In fact, two of three fish that shared AFLP genotypes contained common lake mtDNA haplotypes.

We present two possible hypotheses as to the origin of *G. p. ganzihonensis*. First, the subspecies may have evolved after a population of *G. p. przewalskii* was isolated in the Ganzi River when its lake connection was severed. Second, it is possible that the subspecies was always a distinct riverine population that evolved separately from the lake fish. A 1964 survey found that the Ganzi River lacked a direct surface water connection with Lake Qinghai (Zhu and Wu, 1975). Currently, the lower reaches of the river experience low flows and drain into a marshy area. During heavier flows, the marsh could periodically overflow into the adjacent Haerga River, but this would not allow a direct connection with the lake. It has been proposed that as the lake water levels fell, fish became trapped in the Ganzi River and were forced to adapt to the new environment (Zhu and Wu, 1975). The morphology of the Ganzi River is quite unlike the other rivers in that it originates from a hot spring and contains marshes in the middle and lower reaches. *G. p. ganzihonensis* have fewer gill rakers than *G. p. przewalskii,* a condition thought to have been an adaptation to a different food source in the Ganzi River (Zhu and Wu, 1975).

Although the lake has been regressing for several thousand years, it is not known at what time the direct connection with the Ganzi River was severed (Walker *et al.*, 1996). Some reports suggest that the lake has receded approximately 100 km since the late Pleistocene and historical data indicated that the lake has fallen approximately 12 meters from the late 1800s to the 1990s, with the regression being more rapid in the last 30 years (cf. Walker *et al.*, 1996). Intriguingly, historical records from the Tang Dynasty (608-904 AD) suggest that higher lake levels at the time may have flooded the current lower marsh area possibly allowing a direct connection with the Ganzi River (Zhu and Wu, 1975). If the connection was severed as the lake receded within

the last thousand years, then the morphological changes noted for this subspecies must have occurred quickly, as speculated by Zhu and Wu (1975).

Alternatively, it is possible that the fish did not evolve after being isolated from the lake in recent times, but were always a distinct riverine population with no genetic exchange with the lake fish. In addition to containing gravel areas necessary for spawning, the Ganzi River contains suitable overwintering habitats in the form of deep pools in the marshy areas (Zhu and Wu, 1975); therefore, it is possible that the fish were able to overwinter in the river without ever having to return to the lake. Polymorphisms that existed in the founding population may have been lost due to selection, bottlenecks, or founder effects, resulting in fewer shared haplotypes with the lake fish (Figure 2.2). The relatively high frequency of unique mtDNA haplotypes in the Ganzi, the clustering of some AFLP genotypes with *G. eckloni* genotypes, and the lack of lake-specific genotypes (Figure 2.3) supports an older over a more recent ancestry. However, the presence of lake mtDNA haplotypes within the Ganzi suggest the split from the lake occurred some time after the separation of the Qinghai drainage from the Yellow river drainage.

Applying the criteria of Crandall and colleagues (2000), the two subspecies are likely genetically non-exchangeable, both historically and recently, as significant genetic divergence was detected using both mitochondrial and AFLP markers. The ecological exchangeability between the two subspecies is, however, difficult to assess. Even though the Ganzi River fish occupy a very distinct habitat relative to fish from the other rivers, it is currently unknown if they are able to tolerate the salinity of the lake. Therefore, there is a need for studies similar to that of Wood and colleagues (2007) to determine the physiological effects of moving *G. p. ganzihonensis* from river to lake water. If the Ganzi River fish are unable to acclimate to the salinity of the lake (an indicator of recent ecological non-exchangeability), then, according to

Crandall's criteria, *G. p. przewalskii* and *G. p. ganzihonensis* should be considered as distinct ESU's. If, however, the Ganzi fish are able to reside in the lake and no other evidence exists which suggests a lack of ecological exchangeability, then *G. p. przewalskii* and *G. p. ganzihonensis* should be treated as a single ESU. However, in this case, for the purposes of management and conservation, we would recommend that they be treated as separate Management Units (MUs) (Moritz, 1994).

2.5.3 Lake Qinghai Stock Structure

AMOVA and STRUCTURE results of the AFLP data indicated the presence of a stock structure in Lake Qinghai (Table 2.3 and 2.4, Figure 2.3). Population structuring in the lake was also suggested in separate studies using isozymes (although no statistical analysis was performed) (Meng *et al.*, 2007). Significant differences were also found between fish sampled in the Buha and Shaliu Rivers using microsatellites (Zhang *et al.*, 2009). Population structuring was not detected using mitochondrial analysis, a result that is similar to other studies that combine mtDNA and AFLP data (Giannasi *et al.*, 2001; Sullivan *et al.*, 2004; Egger *et al.*, 2007). It is thought that while mtDNA can detect "larger scale", historical relationships (such as the divergence between *G. p. przewalskii* and *G. p. ganzihonensis* where it is unlikely that gene flow exists), AFLPs are better at detecting finer scale relationships, such as the population structure within the lake (where gene flow between populations may still exist due to straying) (Egger *et al.*, 2007). The formation of distinct populations in anadromous species is primarily due to natal homing (Stewart *et al.*, 2003). Although AFLP analysis indicated that most scaleless carp populations of Lake Qinghai are genetically distinct, the degree of straying or dispersal present among the populations, and the effects of straying on genetic divergence are unknown. It is also unknown if the populations are diverging towards an equilibrium state where the genetically

isolating effects of natal homing are balanced by the homogenizing effects of a low level of migration (straying), or if they are currently at an equilibrium state. Fish species with a strong propensity for natal homing still rely on the mechanisms of dispersal to colonize new habitat or to avoid potentially adverse local conditions (Pascual *et al.*, 1995; Milner *et al.*, 2000; Stewart *et al.*, 2003). Conditions in the Lake Qinghai watershed can be quite severe and large die-offs in spawning runs have been noted due to the rapid changes in water levels that have trapped fish in drying stream beds (Murray, B.W., Pers. Comm.). A possible propensity for straying was also noted by the observation of fish swimming up an ephemeral stream that appeared after a heavy rainfall in June, 2004, but was dry two days later (it is unknown if these individuals were sexually mature at the time). Sexually immature fish have also been observed in the Heima River (Wood *et al.*, 2007), although the reason for their presence is unclear.

Despite the detection of a stock structure within the lake, samples collected from the Quanji and Shaliu Rivers were not found to be significantly different from each other. There may be a significant amount of gene flow between fish in these two tributaries (*i.e.*, more straying than natal homing). Unstable habitats and anthropogenic disturbances (*i.e.* decreased flows due to the installation of weirs and dams) may promote higher levels of straying (Quinn and Dittman, 1990). Both rivers have had recent (since the 1960s) weir and water diversion projects that have impeded access to upstream spawning habitat and have affected water flow at downstream spawning sites. It is also possible that fish from one river may have been extirpated but repopulated by fish from the other river. Again, the isozyme study conducted by Meng and colleagues (2007) suggests that these two populations (from the Quanji and Shaliu Rivers) clustered together.

It is particularly interesting that the fish sampled from two different locations on the Buha River (approximately 50 km from each other) were found to be significantly different, possibly indicating the presence of distinct spawning populations within the river. The evolutionary benefits of natal homing, namely increasing the chances of locating mates or habitats suitable for early life history survival, may result in selective forces that develop within river subpopulations (Garant *et al.*, 2000). The Buha River is the largest tributary of Lake Qinghai (Figure 2.1), contributing approximately half of the total inflow (Walker *et al.*, 1996). It has not been affected by weirs as in the other tributaries, possibly allowing for the continued existence of distinct within-river stocks. Since there have been reports of juveniles overwintering in the upper reaches of the Buha River (Walker and Yang, 1999), it is also possible that a resident population exists in the upper reaches of the Buha. Since very little is known about the life history of the scaleless carp (*e.g.,* how long they reside in the river as juveniles before travelling to the lake), however, this theory is speculative and would need to be confirmed with further studies. The detection of subpopulations within the Buha River also has implications for the management of scaleless carp. It may be possible that similar within-river structuring exists in other spawning tributaries, but went undetected as only one location was sampled in each of these rivers.

The lack of mitochondrial divergence between the different sampling locations indicates genetic exchangeability on a historical level, whereas significant divergence at AFLP markers may indicate recent genetic non-exchangeability or natal homing with low levels of straying among populations. Ecological exchangeability refers to the ability of individuals to occupy the same niche, due to shared life history traits and morphology. It seems likely that the fish of Lake Qinghai would be ecologically exchangeable, although further studies are required for this to be accurately determined. Applying the criteria of Crandall and colleagues (2000), *G. p.*

przewalskii should be treated as a single ESU for conservation purposes. However, exchangeability of the populations should not be used to promote homogenization of fish spawning in the different rivers (Crandall *et al.*, 2000). As some degree of genetic divergence was detected in this study, we recommend that the spawning groups be treated as separate MUs.

It is clear from this research that much more information is needed if scaleless carp stocks are to be managed properly. Although the genetic data in this study indicate a significant level of homing, the degree of straying is unknown as are the exact spawning locations. Unfortunately, the lack of clustering of genetic groups with species or populations precludes the use of STRUCTURE type approaches with the AFLP marker systems to detect migrants. Future studies should therefore include tagging, so that the fish may be accurately tracked to their spawning locations. Straying rates can then be determined by combining tagging data with genetic analysis so that this unique and possibly threatened population can be protected and managed.

2.6 ACKNOWLEDGEMENTS

This research was supported by a National Natural Science Foundation of China Excellent Young Investigator award (30128016) to YXW and JZD, by the Zhejiang U. Overseas Researcher Program for BWM and YXW, a Tang Yongqian Foundation Fellowship at Zhejiang U. to CMW, and by Natural Sciences and Engineering Research Council Canada Discovery Grants awarded to BWM, CJB, JGR, and CMW. We wish to thank the Fishery and Aquaculture Bureau of Qinghai Province for providing the technical and logistical assistance required for this research expedition. We also thank Yaqing Liu of the Northwest High Plateau Institute of Biology for providing excellent logistical assistance; graduate students Shen Wang and Yibin Cao, Zhejiang U., Joe Rogers, McMaster U., Barry Madison, Queen's U, and Jeffrey Semple, U.

of Waterloo, for their field assistance; Mark Thompson of the UNBC Genetics Facility for laboratory assistance; and Mark Shrimpton and anonymous reviewers for helpful criticism.

3 PRELIMINARY ANALYSIS OF MAJOR HISTOCOMPATIBILITY COMPLEX CLASS IIB SEQUENCE VARIATION IN SCALELESS CARP (GYMNOCYPRIS PRZEWALSKII) OF LAKE QINGHAI, CHINA

3.1 ABSTRACT

The scaleless carp, *Gymnocypris przewalskii przewalskii*, is a threatened fish species endemic to Lake Qinghai, a high pH, high elevation saline lake located on the Qinghai-Tibetan Plateau of western China. Previous studies indicate that *G. p. przewalskii*, its most recent ancestor, *G. eckloni*, and its subspecies, *G. przewalskii ganzihonensis*, are genetically distinct at mtDNA alleles and possibly represent separate Evolutionary Significant Units (ESUs), though additional studies are needed for confirmation. Furthermore, AFLP analysis found significant differences between fish sampled in most tributaries, indicating that the spawning groups should be treated as distinct Management Units (MUs). Genes of the Major Histocompatibility Complex (*Mhc*) are highly polymorphic and under strong positive selection; therefore, sequence variation at these genes may be used to help further investigate the genetic relationships. In this study, two different locus groups were identified (*Gypr-DAB1* and *Gypr-DAB3*), which were similar in the coding region, but highly variable in the first intron. Phylogenetic analyses of exonic and intronic sequences from the *Gypr-DAB1* locus showed that the scaleless carp sequences are more related to each other than to the *Barbus* or common carp sequences, while exonic sequences from the *DAB3* locus do not indicate species-specific clustering, though relationships were not well resolved due to low bootstrap support. Derived amino acid alignments revealed that many of the residues important for mammalian *Mhc* function are conserved in scaleless carp. Wu-Kabat variability plots showed that the most variable amino acid positions correspond with regions that are known to interact with peptides. Furthermore, tests of selection indicated that positive or diversifying selection ($dN > dS$) is occurring in the

peptide binding regions of the scaleless carp sequences (*Gypr-DAB1*: Z=2.348, p=0.010; *Gypr-DAB3*: Z=2.411, p=0.009), but not in the non-peptide binding regions (*Gypr-DAB1*: Z=0.311, p=1.00; *Gypr-DAB3*: Z=-0.536, p=1.00), as seen in the expressed *Mhc* genes of many other taxa. Although much information was acquired from this study, more is needed to fully characterize *Mhc* in scaleless carp.

3.2 INTRODUCTION

Scaleless carp (*Gymnocypris przewalskii*) is a threatened fish species[4] endemic to Lake Qinghai, a saltwater lake located within the Tibetan Plateau in the Qinghai province of western China at an elevation of 3200 m. Although it is the largest lake in China (surface area 4600 km^2), few fish species reside in Lake Qinghai, in part due to extreme conditions (high pH, saline) (Zheng, 1997). *G. przewalskii* were heavily impacted by overfishing between the 1950s and 1990s, and stocks were not able to recover as the fish are very slow growing, taking 7 years to reach sexual maturity (Walker *et al.*, 1996).

Scaleless carp stocks have also been impacted by dams and agricultural diversions on freshwater tributaries, resulting in a significant loss of spawning habitat. In addition, water levels have fallen at a rate of approximately 10-12 cm per year over the last 50 years, which has increased salinity levels in the lake (Walker and Yang, 1999). Salinity increases have occurred since 1978, and current data indicates that this increase is accelerating (Wang *et al.*, 2003; Yang *et al.*, 2005; Wood *et al.*, 2007). Although scaleless carp are currently able to physiologically tolerate the conditions of the lake (Wood *et al.*, 2007), experiments have shown that an increase in salinity levels may be lethal (Brauner, C.J., unpublished data).

A comparison of D loop mtDNA sequences between *G. przewalskii* and its ancestral species, *Gymnocypris eckloni*, found no sharing of haplotypes but incomplete lineage sorting, indicating a close evolutionary relationship between the two species. In addition, analysis of sequence variation of the control region of the mitochondrial genome also found that a subspecies (*G. przewaslkii ganzihonensis*), located within the Ganzi River, a tributary that lacks

[4] currently listed as "Grade B" on a list of priority species by the China Biodiversity Conservation Action Plan (Zhan, 1994).

direct surface connection to the lake, is genetically distinct from the lake fish, and may represent a separate Evolutionary Significant Unit (ESU), although additional studies are needed for confirmation. Furthermore, Amplified Fragment Length Polymorphism (AFLP) analysis of scaleless carp populations has determined that fish sampled in most tributaries are genetically distinct and may be at the early stages of lineage divergence. Fish from different tributaries should be treated as separate Management Units for conservation purposes (Chapter 2). As such, a further loss of spawning habitat may impact the genetic diversity of the entire species; however more information is needed if scaleless carp populations are to be managed properly.

Genes of the Major Histocompatibility Complex (*Mhc*) have been shown to be highly polymorphic and under strong positive selection; therefore, isolation of these genes in scaleless carp may assist in exploring the role of selective pressures in shaping the stock structure of the lake (General Introduction). Isolation of these genes may also further assess the genetic differences between *G. przewalskii, G. p. ganzihonensis,* and *G. eckloni*. However, genetic variation at *Mhc* loci in scaleless carp and related species and subspecies has not yet been described. Therefore, this study has three objectives: 1) to investigate *Mhc* Class IIB sequence variation in scaleless carp, 2) to determine the phylogenetic relationship of *Mhc* sequences obtained in this study in comparison with closely related teleosts, and 3) to examine the functional significance of *Mhc* genes in scaleless carp. These objectives were achieved by first examining one sample in detail, followed by a screening of sequence variants in several samples. Information from this examination may allow *Mhc* genes to be used as part of a genetic marker system in order to identify discrete populations or stocks of fish within Lake Qinghai.

3.3 METHODS

3.3.1 Sample Collection

Fish were collected from Lake Qinghai (2002), as well as from the Heima (2004), Buha (2004), and Ganzi Rivers (2005) (Chapter 2). A range of tissue samples, including head kidney, white muscle, liver, and gills, were taken from lethally sampled fish collected in the Buha River and preserved in RNAlater® (Qiagen, Valencia, CA). Samples were kept in a chilled cooler or refrigerator while in the field (no more than 3 weeks) and then transferred to a -30°C freezer until RNA extraction could be performed.

3.3.2 cDNA

Not only is it more time and cost effective to type *Mhc* loci from genomic DNA (rather than cDNA), it is also the preferred method when the amount of lethal sampling should be minimized (for instance, when examining a threatened fish species). However, the amplification of only genomic DNA does not allow one to determine whether functional genes or pseudogenes were isolated. Therefore, expressed *Mhc* genes were also isolated in this study.

Total RNA was extracted from fifteen head kidney samples[5] using the RNeasy® Mini Kit (Qiagen). Prior to RNA extraction, tissues were homogenized using the QiaShredder (Qiagen). The optional on-column DNase digestion with the RNase-free DNase Set (Qiagen) was also performed to reduce DNA contamination in RNA samples. RNA quality and quantity were determined using spectrophotometry with a quartz cuvette per manufacturer's recommendations.

First strand cDNA synthesis was carried out using the Thermoscript™ RT-PCR (reverse transcriptase PCR) System (Invitrogen, Carlsbad, CA). This system is a two step process, the

[5] Twelve samples originated from fish collected from the Buha River in 2004. The three remaining samples originated from lake fish collected in 2002.

first of which uses an Oligo $(dt)_{20}$ primer, followed by amplification with gene specific primers. Primers used to investigate the *Mhc* Class IIB genes in the common carp and other cyprinid species were initially used to target expressed *Mhc* genes in scaleless carp samples. Forward primers targeted the end of the first exon, while the reverse primer targeted the end of the second exon, which codes for the variable peptide binding region (Figure 3.1). The following 8 primer combinations were initially used in this study: OL93-139, OL93-140; OL93-139, OL94-23; Ex1DAB4, OL94-23; OL92-139, OL93-23; Ex1DAB4, DAB04s; Ex1DAB4, DAB06s; DXA-fw, DXA-rev; and OL93-139/CycaDABex2-rev (Table 3.1). Their abilities to amplify *Mhc* genes in scaleless carp were tested on cDNA samples.

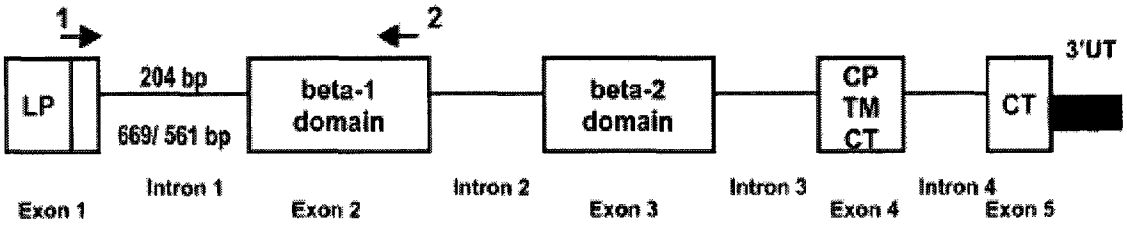


Figure 3.1: Schematic drawing outlining the basic structure of *Mhc* Class IIB genes in *C. carpio* [from (Rakus *et al.*, 2003)]. LP - leader peptide, CP – connecting peptide, TM – transmembrane region, CT – cytoplasmic domain, 3'UT – 3' untranslated region.

Table 3.1: *Mhc* primer sequences.

Primer Name	Primer Sequence	Reference
OL93-139	5'-CTGATGCTGTCTGCTTTCACTGGAGCA-3'	(van Erp *et al.*, 1996)
OL93-140	5'-CTGTTTTATCACGGATCGCTGACTG-3'	(van Erp *et al.*, 1996)
OL94-23	5'-GATTTGAGCATTATGTTTGCA-3'	(van Erp *et al.*, 1996)
Ex1DAB4	5'-ATGCTGTCTGCATTTACTGGAACAG-3'	(van Erp *et al.*, 1996)
OL92-139	5'-CTGATGCTGTCTGCTTTCACTGGAGGAG-3'	(Dixon *et al.*, 1996)
OL93-23	5'-CGATTTGAGCATTATGTTTGCAG-3'	(Dixon *et al.*, 1996)
DAB04s	5'-CTCTGCTGCAGTTCTGCC-3'	(van Erp *et al.*, 1996)
DAB06s	5'-TGTCCACTGAAGTTTTCAGA-3'	(van Erp *et al.*, 1996)
DXA-fw	5'-GCTCAAGCTGAGCACAGGG-3'	(van Erp *et al.*, 1996)
DXA-rev	5'-CTCTTCTGGAGAGTTGTATGC-3'	(van Erp *et al.*, 1996)
CycaDABex2-rev	5'-GAGTCAGCGATCCGTGATAAAACA G-3'	(Rodrigues *et al.*, 1995)

All cDNA samples were tested with each of the 8 primer pairs at each of three annealing temperatures (48°C, 52°C or 55°C). One μL of cDNA was added to 24 μL of master mix containing the following: IX PCR Buffer (Invitrogen, Calsbad, CA), 200 μM of each dNTP, 2 mM of $MgCl_2$, 0.2 μM of each primer, and 1 unit of Platinum TAQ DNA polymerase (Invitrogen). Amplifications were carried out on a DYAD or PTC-100 thermal cycler (MJ Research, Inc., Waltham, MA). The thermal cycle was 96°C for 2 min, 48°C, 52°C or 55°C for 1 minute, and 72°C for 2 minutes, followed by 39 cycles of 94°C for 30 seconds, 48°C, 52°C, or 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds. The final extension phase was 72°C for 6 minutes and 30 seconds.

3.3.3 Genomic DNA

Genomic DNA was extracted from caudal fin clip samples using a standard phenol/chloroform method as in Chapter 2. Genomic DNA was isolated to determine whether functional genes can be detected in genomic DNA. In particular, a comparison of genomic and cDNA sequences was performed to ensure that the genomic loci under investigation are expressed. This was done by analyzing both cDNA and genomic DNA from two individuals.

One µL of genomic DNA (50 ng/µL) was added to 24 µL of master mix containing 1X PCR Buffer (Invitrogen), 200 µM of each dNTP, 2 mM of $MgCl_2$, 0.2 µM of each primer, and 1 unit of Platinum *Taq* DNA polymerase (Invitrogen). Amplifications were carried out as described in the previous paragraph.

3.3.4 Cloning and Plasmid DNA Extraction

PCR products were visualized using agarose gel electrophoresis. Successfully amplified PCR products of the appropriate size (Dixon *et al.*, 1996; van Erp *et al.*, 1996) were excised from the gel and purified using the QIAquick Spin Kit (Qiagen). The concentration of the purified PCR product was measured by running the product on an agarose gel with a mass standard.

The purified PCR product was ligated into a cloning vector using the pGEM®-T Easy Vector System (Promega, Madison, WI). Cloning vectors were inserted into JM109 High Efficiency Competent cells (Promega) using a heat shock treatment. Cultures were incubated in SOC Medium (Invitrogen) for 1.5 hours at 37°C with shaking (approximately 150 rpm).

One hundred µL of each transformation culture were spread over duplicate LB/ampicillin (100 µg/mL)/IPTG(100 mM)/X-Gal (50 mg/mL) plates using sterile techniques. The plates were incubated overnight at 37°C. Successful transformations were indicated by the presence of white

colonies as successful cloning of an insert disrupts the coding sequence of β-galactosidase. Positive colonies were picked off the plate using sterile techniques and inoculated into 3 mL of LB broth with ampicillin (50 μg/ml). Cultures were incubated overnight at 37°C with shaking (approximately 200 rpm).

Plasmid vector DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen). Since the pGEM-T Easy Vector contains *Eco*RI restriction sites that flank the cloning site, a simple digest with this enzyme was used to identify false positives (white colonies that do not contain an insert). Five μL of plasmid DNA were added to 15 μL of a master mix containing 5 units of *Eco*RI and 1X React 3 Bufffer. Solutions were incubated for approximately 3 hours at 37°C. Restriction digests were separated via agarose gel electrophoresis.

3.3.5 Sequence Analysis

Plasmid DNA samples that contained an insert were sequenced on the CEQ 8000 Genetic Analysis System (Beckman Coulter, Inc., Fullerton, CA). Approximately 20 ng of purified DNA were combined with 4 μL of DTCS Quick Start Master Mix, 2 μL of T7 (5'-TAATACGACTCACTATAGGG-3') or SP6 (5'-ATTTAGGTGACACTATAG-3') primers, and enough nuclease-free water to bring the total volume to 10 μL. Sequencing reactions were carried out via 30 cycles of 96°C for 20 seconds, 50°C for 20 seconds, and 96°C for 4 minutes. Sequence fragments were separated using method LFR-1 (denatured: 90°C for 120s; injected: 2.0kV for 15s; and separated: 4.2kV for 85min in a 50°C capillary). Since sequencing a plasmid may be difficult due to supercoiling (Thompson *et al.*, 2008), samples were linearized by digesting with *Sca*I prior to sequencing.

Sequence chromatograms were viewed, edited (*i.e.*, plasmid and primer sequences removed), and aligned as described in Chapter 2. Scaleless carp sequences were compared to other cyprinid sequences obtained using the BLAST feature of GenBank (Table 3.2). Maximum likelihood phylogenetic trees were constructed using the methodology from Chapter 2. To best show the phylogenetic relationships between samples and in comparison to related species, the exons and introns from each of the two loci were examined separately. Wu-Kabat variability plots (Wu and Kabat, 1970) were constructed using the derived amino acid sequences generated in MEGA. The numbers of nonsynonymous (dN) and synonymous (dS) substitutions and Z tests of selection were also computed in MEGA.

Table 3.2: A listing of *Mhc* sequences from closely related species of scaleless carp, used for phylogenetic comparison. *Mhc* locus names are named according to the first two letters of both the genera and species of origin (*'Bain' – Barbus intermedius, 'Babo' – Barbus bocagei; 'Cyca' – Cyprinus carpio*).

Mhc Sequence Name	Accession Number
Bain-DAB1	X93677
Bain-DAB1-06	X93682
Babo-DAB1	X93896
Cyca-DAB1	Z47731
Cyca-DAB2	Z47732
Bain-DAB3	X93895
Babo-DAB3	X93897
Cyca-DAB3	Z47733
Cyca-DAB4	Z95435

3.4 RESULTS

3.4.1 Primer Results

Three out of eight primer pairs were successful in amplifying products from the scaleless carp samples – OL93-139/OL94-23 (55°C annealing temperature), ex1DAB4/OL94-23 (52°C), and OL92-139/OL94-23 (55°C). All products amplified using cDNA yielded product sizes of approximately 280 bp, while genomic DNA samples mostly yielded products approximately 500 bp in size, although some were approximately 900 bp. These larger fragments did not appear to correlate with any one primer pair. Fragments of other sizes were also produced in many samples. A few of these fragments were sequenced, but found to be non-specific.

3.4.2 Sequence Variation

One cDNA sample (14-2) was examined in detail, and fifteen clones from this sample were sequenced (Table 3.3). For the remaining samples, both cDNA and genomic, up to seven clones were sequenced (samples sequenced at one or two clones are indicated by italics in Table 3.3). A large number of samples in this study were examined in collaboration with a fourth year undergraduate applied genetics laboratory course (Biology 425).

Table 3.3: *Gymnocypris przewlaskii Mhc* sequences arranged by sample and locus. Numbers indicate the number of samples that contain that particular locus. g = genomic derived clone, c = cDNA derived clone, $x^{b/c/d/e}$ = 1 bp variant noted from a single clone (to be confirmed). x^{d} variant is located in the intron. Sample numbers in italics were sequenced once or twice.

Sample	*Gypr-DAB1*																*Gypr-DAB3*			
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	01	02	03	04
14-2			2c	13c																
15-6	2g1c																			
10-44														3g	1g	1g				
12-52							4g													
12-53					1g	2g														
QH1-1									2c	1c							1c			
QH1-2	1^{b}c																6c			
QH1-3		2g1c																		
12-42											1g									
12-43								1g												
13-2													2g							
14-1		1^{d}g																		
14-5												1g								
15-8		1g																		
U1-1	1c																			
U1-2	1^{e}c																			
U1-3																				1c
U1-4																			1c	
U1-5																			1^{e}c	
U1-6																		1c		
U1-7																		1g		

3.4.3 cDNA

A total of 10 cDNA sequences were obtained from 11 different samples, although several samples contained single base pair variations (Table 3.3). Resulting sequences were 226 bp, which was the expected size of the amplified section of the second exon of *Mhc* genes in common carp and *Barbus spp* (Dixon *et al.*, 1996). Multiple sequences were observed in individuals for whom more than one clone was analyzed (Table 3.3). Clones with single nucleotide differences (noted in Table 3.3) were not considered new sequences. These may represent real genetic variation or they may represent PCR artifacts; therefore, samples possessing these single base pair variants should be sequenced using additional clones to confirm variation.

3.4.4 Genomic

Twelve genomic sequences, approximately 440 bp in length, the expected size of the *DAB*1/2 locus (intron 1 and exon 2) in common carp and *Barbus spp* (Dixon *et al.*, 1996), were obtained from 11 samples. One 965 bp sequence also resulted from sequencing a larger PCR fragment. This is the expected product size of a second *Mhc* locus, corresponding to *Cyca-DAB3-01* and *Cyca-DAB4-01* in the common carp (van Erp *et al.*, 1996).

Ten genomic sequences were unique to a single sample. Three sequences (*Gypr-DAB1-01, DAB1-02, and DAB3-02*) had matching exonic sequences to the previously identified cDNA sequences. Matching *DAB1-01* and *DAB1-02* exonic sequences were found in samples (15-6 and QH1-3) where both cDNA and genomic DNA were analyzed, *DAB1-02* was also found in two other genomic samples.

A sequence alignment confirmed that the 440 bp genomic PCR product corresponded to *Mhc* loci *DAB1* and *2* in *Cyprinus carpio* and *DAB1* in *Barbus spp* (Figure A3.1). The length of the intron is 215, similar to *C. carpio* (van Erp *et al.*, 1996), although the length of the intron varies in different individuals due to the presence of multiple short indels (6-7 bp). One out of three genomic sequences obtained from individual 10-44 contained a large insertion (*Gypr-DAB1-16g*).

Sequence alignment also confirmed that the 965 bp sequence and several cDNA samples correspond to the *DAB3* locus in cyprinid species (742 bp intron 1 and 223 bp exon 2) (Figure A3.2).

DAB1 and *DAB2* in *C. carpio* were originally thought to be different alleles, but subsequent studies indicated that they do not segregate from each other (van Erp *et al.*, 1996). Hence, it was determined that they represent a pair of closely linked genes. Similarly, another pair of linked genes, *Cyca-DAB3* and *Cyca-DAB4*, were found to segregate from the *DAB1* and *DAB2* genes (van Erp *et al.*, 1996). The two groups were similar in the coding regions, but vastly different in the length of their introns. Later studies involving *Barbus spp.* also found two locus groups named *DAB1* and *DAB3* (rather than *DAB1* and *2*, and *DAB3* and *4*). These two groups were presumed to be two distinct loci, which may or may not contain several different genes (Dixon *et al.*, 1996). For the purposes of this study, all scaleless carp sequences were named following the same nomenclature rules as in previous studies, with the first locus group named *Gypr-DAB1* and the second group, *Gypr-DAB3*.

3.4.5 Phylogenetic Analysis

A phylogenetic comparison of *DAB1* exon 2 sequences (obtained from cDNA samples as well as the exon 2 portion of genomic samples) with *C. carpio* and *Barbus spp.* shows species-specific clustering with high bootstrap values (Figure 3.2). There are four groups of samples that cluster together with bootstrap values greater than 50%; however the relationships between the majority of scaleless carp sequences are not well supported by high bootstrap values. It is interesting to note that, of the three sequences originating from sample G10-44, two (*Gypr-DAB1-14* and *DAB1-15*) are in the same group, while the third (*Gypr-DAB1-16*) is separated into a different group. This may indicate that sequences 14 and 15 represent different alleles of the same gene.

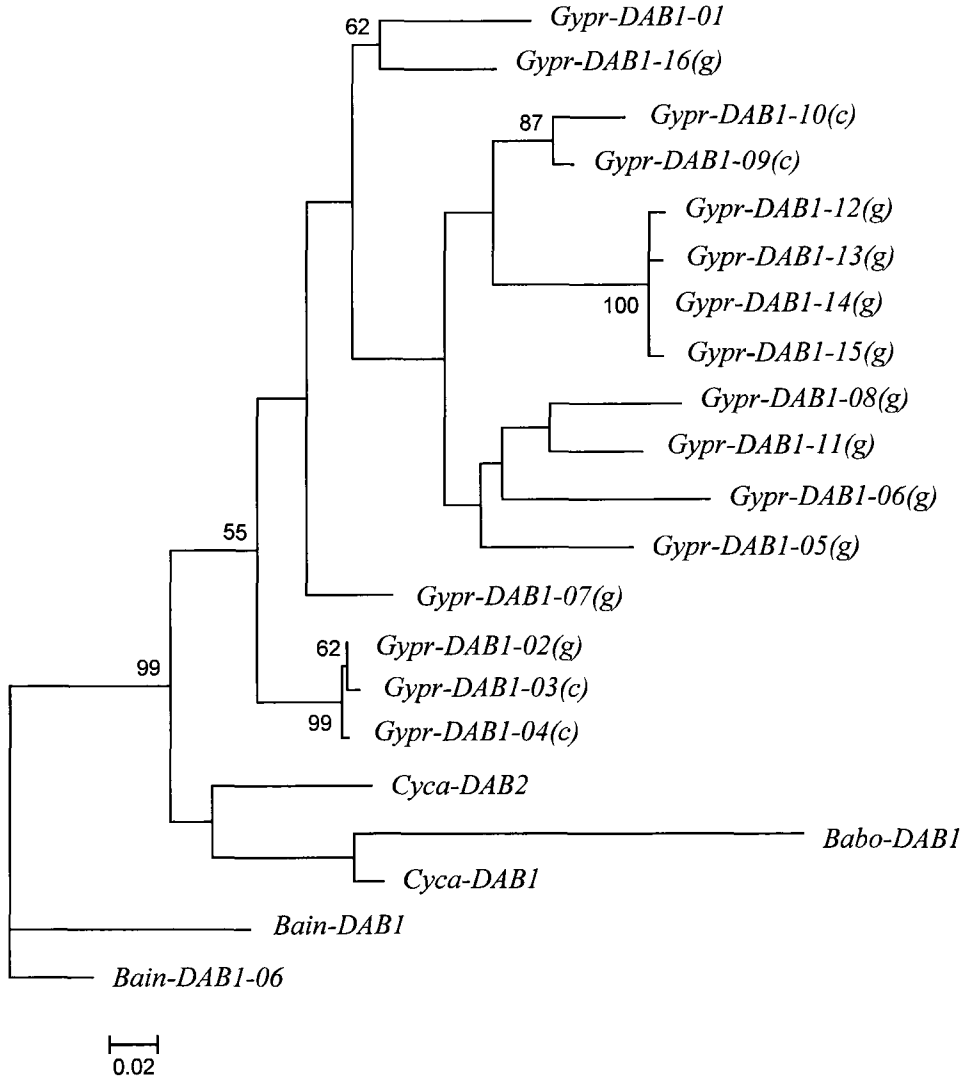


Figure 3.2: Maximum likelihood tree constructed using exon 2 sequences of *Mhc-DAB1* locus group in *G. przewalskii* (*Gypr*), *C. carpio* (*Cyca*), and *Barbus spp.* (*Bain* and *Babo*). The best fit model of nucleotide substitutions identified using the Bayesian Information Criterion (BIC) implemented in jModeltest 0.1.1 (Posada, 2008), JC69 with unequal base frequencies, was used by program PHYML (Guindon and Gascuel, 2003) to estimate the above phylogeny and calculate bootstrap values. A model-averaged phylogeny (Posada, 2008) using relative model weights based on either the BIC or Akaike Information Criterion is identical in topology to the tree shown with most nodes having a model support of 1.0. Bootstrap values over 50% are shown.

Unlike the *DAB1* locus, exonic sequences from the *DAB3* locus do not cluster together in a species-specific manner (Figure 3.3). *Gypr-DAB3-01* is interspersed between sequences from the comparative species, and *DAB3-02* appears to be more related to the *B. intermedius* sequences than to the other scaleless carp sequences. As most bootstrap values are below 50%, it is difficult to precisely determine whether or not there are clusters or groups of samples that could indicate different alleles of the same gene. However, given that sequences 1 and 2 appear quite distant from sequences 3 and 4, it seems likely that there may be more than one cluster.

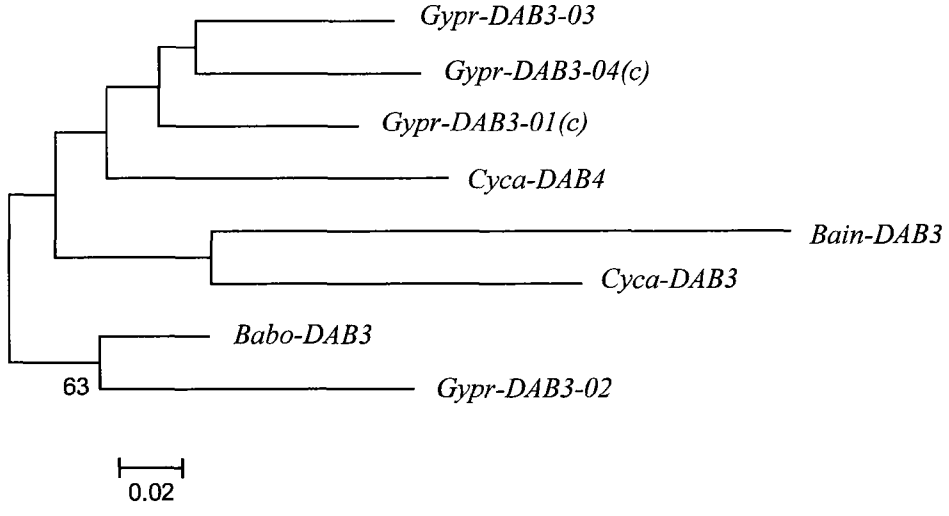


Figure 3.3: Maximum likelihood tree constructed using exon 2 sequences of *Mhc-DAB3* locus group in *G. przewalskii* (*Gypr*), *C. carpio* (*Cyca*), and *Barbus spp.* (*Bain* and *Babo*). The best fit model of nucleotide substitutions identified using the Bayesian Information Criterion (BIC) implemented in jModeltest 0.1.1 (Posada, 2008), JC69 with unequal base frequencies, was used by program PHYML (Guindon and Gascuel, 2003) to estimate the above phylogeny and calculate bootstrap values. A model-averaged phylogeny (Posada, 2008) using relative model weights based on either the BIC or Akaike Information Criterion is identical in topology to the tree shown with all nodes having a model support of 1.0. Bootstrap values over 50% are shown.

DAB1 intronic sequences reveal the clustering of all common carp and all *Barbus* species, while the relationships among the scaleless carp sequences are generally not well supported by

high bootstrap values (Figure 3.4). However, as in exon 2, two sequences from sample 10-44 cluster together, while the third is located on a different branch, again, possibly indicating different alleles. Sequences 1 and 2 could also represent a third group, as they cluster together away from the other sequences with a bootstrap value greater than 50%. Sequences from sample 12-53 (*Gypr-DAB1-05g* and *06g*) appear to be divergent, which was not seen when examining the exonic sequences. A phylogenetic analysis of *DAB3* intronic sequences has been included for consistency, although only one complete sequence from *G. przewalskii* was available (Figure 3.5).

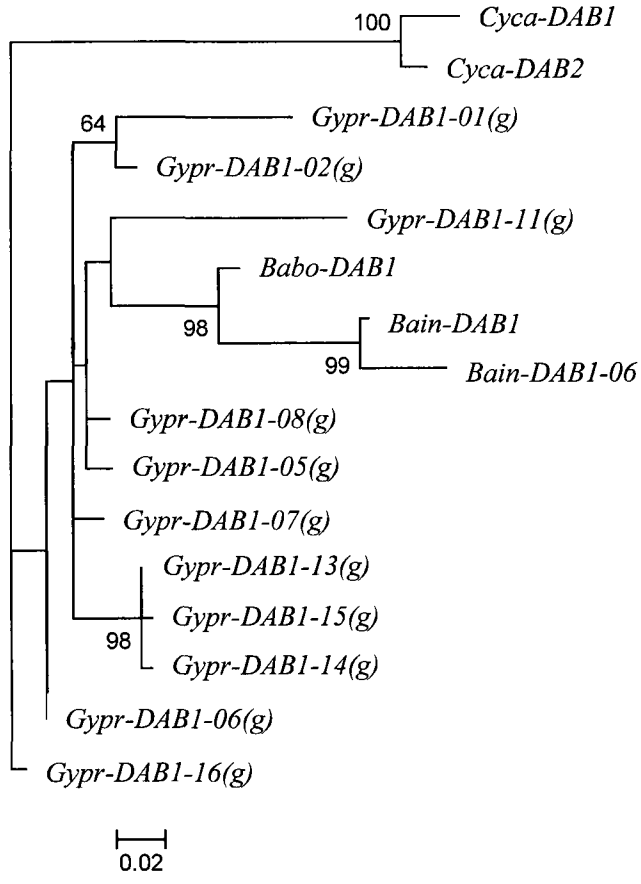


Figure 3.4: Maximum likelihood tree constructed using intron 1 sequences of *Mhc-DAB1* locus group in *G. przewalskii* (*Gypr*), *C. carpio* (*Cyca*), and *Barbus spp.* (*Bain* and *Babo*). The best fit model of nucleotide substitutions identified using the Bayesian Information Criterion (BIC) implemented in jModeltest 0.1.1 (Posada, 2008), HKY with unequal base frequencies, was used by program PHYML (Guindon and Gascuel, 2003) to estimate the above phylogeny and calculate bootstrap values. A model-averaged phylogeny (Posada, 2008) using relative model weights based on either the BIC or Akaike Information Criterion is identical in topology to the tree shown with most nodes having a model support of 1.0. Bootstrap values over 50% are shown.

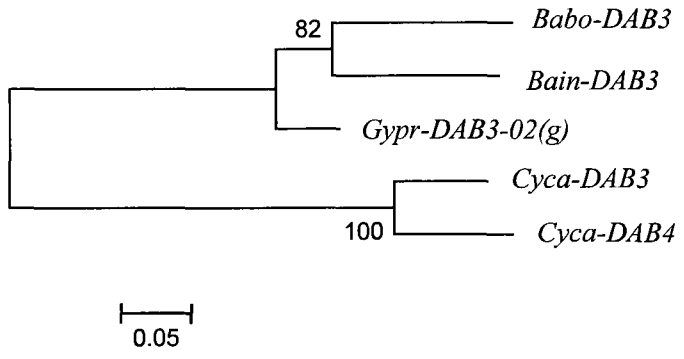


Figure 3.5: Maximum likelihood tree constructed using intron 1 sequences of *Mhc-DAB1* locus group in *G. przewalskii* (*Gypr*), *C. carpio* (*Cyca*), and *Barbus spp.* (*Bain* and *Babo*). The best fit model of nucleotide substitutions identified using the Bayesian Information Criterion (BIC) implemented in jModeltest 0.1.1 (Posada, 2008), K80 with unequal base frequencies, was used by program PHYML (Guindon and Gascuel, 2003) to estimate the above phylogeny and calculate bootstrap values. A model-averaged phylogeny (Posada, 2008) using relative model weights based on either the BIC or Akaike Information Criterion is identical in topology to the tree shown with all nodes having a model support of 1.0. Bootstrap values over 50% are shown.

3.4.6 Expression of *Mhc* genes

Both genomic and cDNA sequences were isolated from only two individuals (15-6 and QH1-3 in *DAB1*). The exonic sequences from each sequence pair were identical, indicating that these two genomic sequences are expressed in these individuals.

Alignment of the derived amino acid sequences from exon 2 in both the *Gypr-DAB1* and *Gypr-DAB3* loci again confirmed that the sequences obtained correspond to *Cyca-DAB, Bain-DAB,* and *Babo-DAB* genes (Figure 3.6). As in Dixon *et al.* (1996), many of the scaleless carp amino acid residues important for mammalian *Mhc* functioning are conserved. The cysteine at positions 9, important for the formation of disulphide bridges that stabilize the peptide binding groove; glycines at positions 40 and 49 (important to proteins' secondary structure); and arginine or lycine residues at position 67 are conserved. There is also an N-linked glycosylation site at positions 36 to 38, which is conserved in cyprinid and salmonid species, although not in

mammals (Dixon *et al.*, 1996). Four *Gypr*-DAB1 amino acid sequences also contain an additional putative N-linked glycosylation site at positions 58 to 60, also found in some *Barbus* sequences.

```
Gypr-DAB1-11(g)   YYHSTWSRCIHSSRDLSDMVYIDNYIFNKDVDIQFNSTVGEYVGYTEHGVYNAQLWNNDTNLLQQERADVER
Gypr-DAB1-16(g)   ..N.R.AK........T.I.FV.........VV.............L..KS..S..S.P.R......Q..T
Gypr-DAB1-14(g)   ......I..F........I.F..........Y.........................K.P.....M..Q..T
Gypr-DAB1-15(g)   ......I..F........I.F..........Y.........................K.P.....M..Q..T
Gypr-DAB1-08(g)   ..Y.R.G.........T..............YV.............AY..K...F...N.............
Gypr-DAB1-07(g)   ..NYV.NK............FV.........F..............L............P.R......E...
Gypr-DAB1-05(g)   ....R.T.........T...M..........YV.............AY......A.....IG......E...
Gypr-DAB1-06(g)   ..S.FRR.............FVA.S......Y.............A..................T..E..T
Gypr-DAB1-13(g)   ......I..F........I.F..........Y..........A..............K.P.....M..Q..T
Gypr-DAB1-02d(g)  ..S.Y.T.........T.I.FV.........F.........F.....L.........K.P.........Q..T
Gypr-DAB1-03(c)   ..S.Y.T.........T.I.FV.........F.........F.....L.........K.P.........Q..T
Gypr-DAB1-04(c)   ..S.Y.T.........T.I.FV.........F.........F.....L.........K.P.........Q..T
Gypr-DAB1-12(c)   ......I..F........I.F..........Y.............H.............K.P.....M..QA.T
Gypr-DAB1-01(g/c) ..N.R.TK........T.....Q.......MVV.........F.....L..H.........P.R...V..QAGT
Gypr-DAB1-02(g)   ..S.Y.T.........T.I.FV.........F.........F.....L.........K.P.........Q..T
Gypr-DAB1-09(c)   ....R.T...Y.....T.I.F..........Y..............A........R.S.P.R...M..Q..T
Gypr-DAB1-10(c)   ..Y.F...........T..............Y..............A........R.S.P.R...M..Q..T
Gypr-DAB1-01b(c)  ..Y.R.TK........T.....Q.......MVV.........F.....L..H.........P.R...V..QAGT
Gypr-DAB1-01c(c)  ..N.R.TK........T...C.Q.......MVV.........F.....L..H.........P.R...V..QAGT

Bain-DAB1-06      ..F.Y.NK...G...F.....VK.I......F.........F.....L.....E...K.P.....W..Q..T
Bain-DAB1         ..F.V.TN.F.G...F.....VA.W......F.........F.....L..H..ES..K.P.R...T..Q..T
Babo-DAB1         ..F.M..............F.N.........F.........YW.....R..H...RF....FF..R..TL...
Cyca-DAB1         ....W.TK.......F..............Y..............AL.....ERF.K.P.I......Q...
Cyca-DAB2         ..R.R.NK.......F....FV.........H.........F....AL..H..E...K..TG......Q.DS

Gypr-DAB3-02(c)   ..SYQM.E.VY.TS.YR....LLS.S..QA..L.Y......F..S.AQ..KY.ENF.KNPAIM..LK.N.D.
Gypr-DAB3-03(c)   ..TY.M.E.VY.TS.Y....FLES.S..EV..LL...S..KF....AE..KF.ENF.K.QAFI...K.A.DT
Gypr-DAB3-03e(c)  ..TY.M.E.VY.TS.Y....FLES.S..EV..LL...S..KF....AE..KF.ENFSK.QAFI...K.A.DT
Gypr-DAB3-04(c)   ..EY.SYD.VY.TS.Y....FLALFS...V..V....S..KF.....E..KY.EN..K.QAII..LK.N.D.
Gypr-DAB3-01(c)   ..EYIEFE.VY.TS.Y....FLTS.S...V..V.Y.....KF.....E.LKY.ENF.K.QAFM..QK.A.DT
Gypr-DAB3-02f(g)  ..SYQM.E.VY.TS.YR....LLS.S..QA..L.Y......F..S.AQ..KY.ENF.KNPAIM..LK.N.D.

Bain-DAB3         ..EEWMPE.VY.A..YR....LVS.F..EAMHA.Y.....KF....GV..KD.EY...NQDYM.GLKG....
Babo-DAB3         ..SYGMPE.VY.TS.Y.....LVS.S...V..........KF.....Q..KF.ENF.KNPAIM..LKGE.D.
Cyca-DAB3         ..EYMTFD.VY.AS.Y.....LLS.S..QGL.A....S..KF.....K.MKD.EY...NPAE...R..E.DT
Cyca-DAB4         ..EY.MYE.VY.TS.Y.....LVSLS..QV..V.C..SAMKC.....E..KY.ENF.K.PSV..DLKTS.DT
```

Figure 3.6: Derived amino acid alignment of exon 2 from DAB1 and DAB3 locus groups in *G. przewalskii* and comparative species. Variable sites in clones for which there is only one sequence will need to be confirmed. "G" represents samples of genomic origin. "C" represents samples of cDNA origin.

The most variable amino acid positions, as shown in the Wu-Kabat variability plots (Figures 3.7 and 3.8) correspond with the regions of human *Mhc* proteins known to interact with peptides (Brown *et al.*, 1993; Dixon *et al.*, 1996), although this trend is seen more clearly in the *Gypr-DAB1* loci (Figure 3.7) versus the *Gypr-DAB3* loci (Figure 3.8).

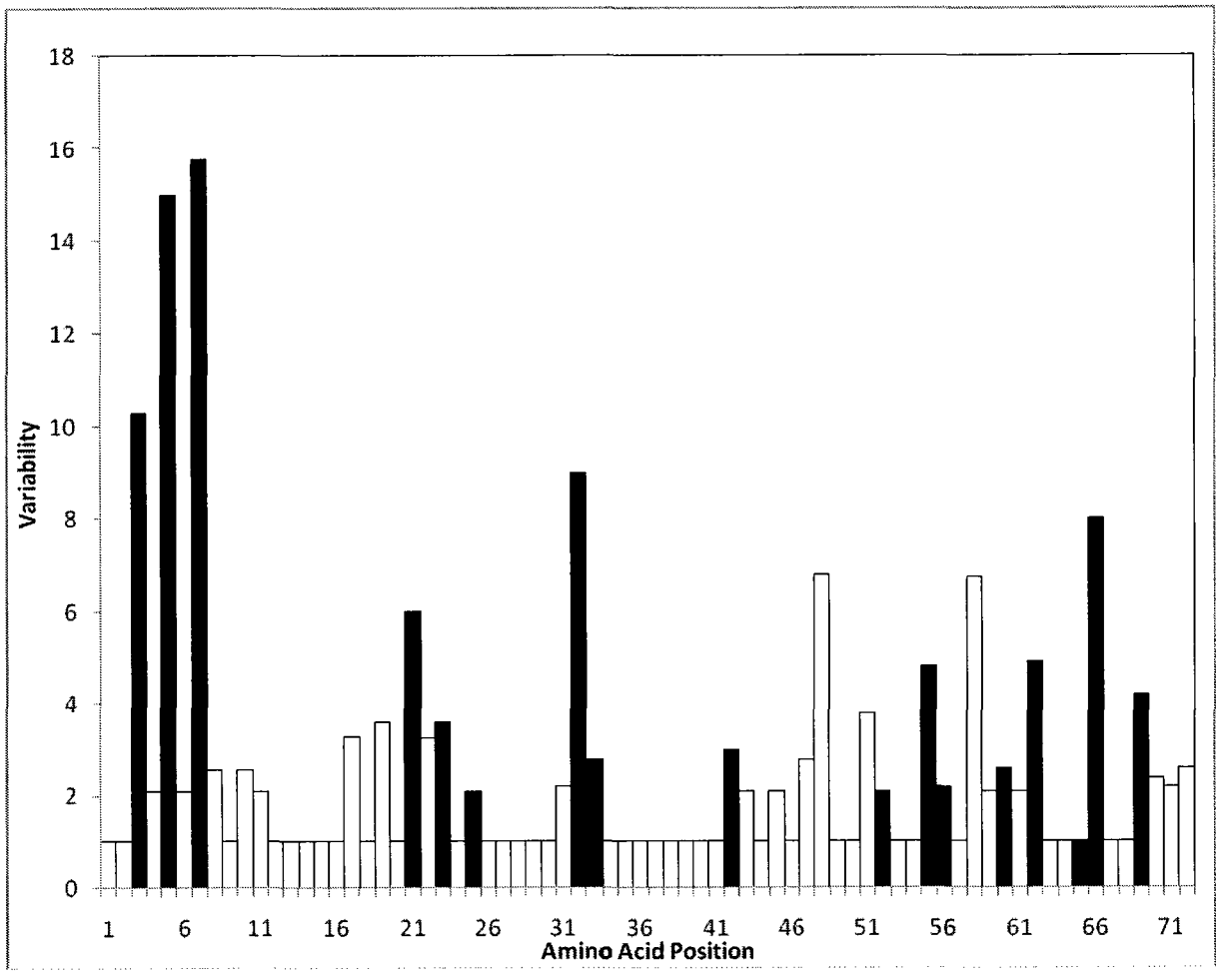


Figure 3.7: Wu-Kabat variability plot (Wu and Kabat, 1970) of derived amino acid sequences of *Mhc* Class IIB (DAB1) in *Gymnocypris przewalskii* (black bars indicate amino acids known to interact with peptides in human *Mhc* Class II beta molecules) (from Brown *et al.*, 1993; Dixon *et al.*, 1996).

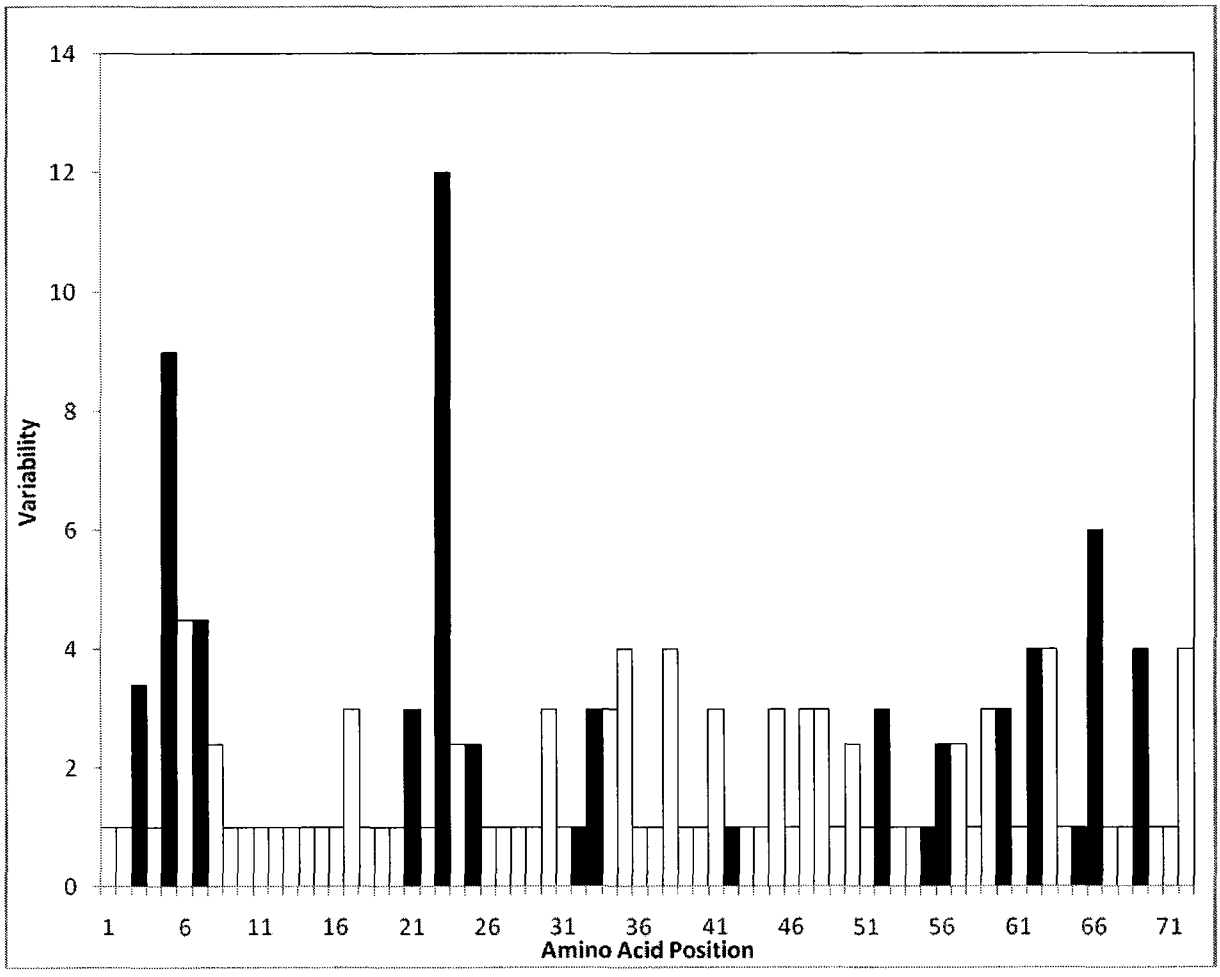


Figure 3.8: Wu-Kabat variability plot (Wu and Kabat, 1970) of derived amino acid sequences of *Mhc* Class IIB (DAB3) in *Gymnocypris przewalskii* (black bars indicate amino acids known to interact with peptides in human *Mhc* Class II beta molecules) (from Brown *et al.,* 1993; Dixon *et al.*, 1996).

3.4.7 Test of Selection

In order to determine whether selective pressures acting on a *Mhc* gene coincide with its function, it is necessary to compare codons of the peptide binding region to codons of non-peptide binding regions. The codons that corresponded to the 17 amino acid residues identified by black boxes in the Wu-Kabat plots (Figures 3.7 and 3.8) were designated the peptide binding codons (Dixon *et al.*, 1996).

The peptide binding regions in both the *Gypr-DAB1* and *Gypr-DAB3* loci have higher rates of nonsynonymous substitutions than synonymous substitutions (Tables 3.4 and 3.5). Z tests indicated that positive selection (dN>dS) is occurring in the peptide binding regions of both loci (Z=2.348, p=0.010 and Z=2.411, p=0.009, respectively), but not in the codons of non-peptide binding regions (Z=0.311, p=1.00 and Z=-0.536, p=1.00, respectively).

Table 3.4: The average number of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) in peptide binding codons (PBR) and non-peptide binding codons (non-PBR) in *Gymnocypris pzewalskii* DAB1 exon 2 sequences.

	Nonsynonymous substitutions per nonsynonymous site (dN)	**Synonymous substitutions per synonymous site (dS)**
PBR	0.180±0.034	0.068±0.054
Non-PBR	0.050±0.015	0.059±0.028

Table 3.5: The average number of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) in peptide binding codons (PBR) and non-peptide binding codons (non-PBR) in *Gymnocypris przewalskii* DAB3 exon 2 sequences.

	Nonsynonymous substitutions per nonsynonymous site (dN)	**Synonymous substitutions per synonymous site (dS)**
PBR	0.291±0.061	0.130±0.055
Non-PBR	0.075±0.016	0.093±0.034

3.5 DISCUSSION

3.5.1 Results Summary

In this study, two different approaches were taken to characterize *Mhc* variation in scaleless carp: examining one sample in detail, followed by a screening of several different samples to identify variant sequences. First, the cDNA of one individual was examined in detail

(15 clones), resulting in two sequences. These sequences are from the same locus (DAB1) and appear to be closely related (Figure 3.4). Second, a smaller number of clones (up to 7) from several different samples were sequenced. As a result, two different locus groups were identified (*Gypr-DAB1* and *Gypr-DAB3*), based on sequence alignment and previous knowledge of *Mhc* Class II beta genes in common carp and *Barbus spp.* (Dixon *et al.*, 1996; Kruiswijk *et al.*, 2004). While the two locus groups are similar in the coding region, the intron 1 regions are variable in both length and nucleotide composition, as seen in common carp and *Barbus spp.* (Dixon *et al.*, 1996; van Erp *et al.*, 1996). The *DAB1* locus group in the scaleless carp may contain more than one locus, as indicated by the clustering of sequences (Figures 3.2 and 3.4). For instance, two out of three sequences from a single sample (10-44) were grouped together in both the exon 2 and intron 1 phylogenies, while the third, which contains an indel in the intron, was located on a separate branch.

Mhc genes in scaleless carp are highly variable, and, in terms of function and selection, behave in a similar fashion as in other teleosts. *DAB1* was the most frequently expressed locus detected in this study, as it appeared in the majority of samples. *Gypr-DAB3*, however, seemed to appear quite randomly (not primer specific), the reason for which is unknown; however it should be noted that all but one of the *DAB3* sequences were isolated from cDNA (head kidney), with only one originating from genomic DNA (fin clip). Expression of one locus over the other may be tissue specific (Rodrigues *et al.*, 1998b), timing specific (in terms of development or life stage) (Rodrigues *et al.*, 1998b), or even temperature dependant (Rodrigues *et al.*, 1998a). Although a range of different tissues was initially collected from lethally sampled fish, nucleic acids in this study were only isolated from head kidney (generally one of the richest sources of

cDNA) or fin clips. Future works could therefore include isolating DNA from these other tissues to determine whether there is any pattern in the expression of *DAB1* or *DAB3* loci.

3.5.2 Phylogenetic Analysis

Phylogenetic analyses of exonic and intronic sequences from the *Gypr-DAB1* locus showed that the scaleless carp sequences are more closely related to each other than to the *Barbus* or common carp sequences, while exonic sequences from the *DAB3* locus do not indicate species-specific clustering; however relationships within the scaleless carp are not well resolved. Trans-species polymorphism (when clustering of sequences is not species-specific) is commonly found in *Mhc* genes (both *DAB1* and *3* loci) of other teleost groups (Dixon *et al.*, 1996; Kruiswijk *et al.*, 2005). A lack of trans-species polymorphism may result from shared allelic lineages being lost in bottlenecks, as seen in salmonid *Mhc* genes (Miller and Withler, 1996; Miller *et al.*, 1997). Cytochrome b data indicates that the subfamily Schizothoracine, of which *Gymnocypris* is a member, may have arisen from other cyprinids approximately 10 mya (He *et al.*, 2004). This is within the time frame observed for the maintenance of polymorphisms via balancing selection in primates (Klein, 1987; Nei and Hughes, 1991). However, the absence of trans-species polymorphism in scaleless carp *DAB1* sequences may indicate that much of its *Mhc* diversity has arisen since its divergence with its comparison species (common carp and *Barbus spp.*). Genetic evidence indicates that the Schizothoracine fishes may have undergone a bottleneck after their divergence from other members of the Cyprinidae due to geological changes during the formation of the Tibetan Plateau (He *et al.*, 2004).

3.5.3 Functional Significance

The amplification of genomic DNA does not allow for the discrimination of functional genes from pseudogenes. As such, Wu-Kabat variability plots were constructed from derived

amino acid sequences (Figures 3.7 and 3.8). Polymorphisms in functional genes should be concentrated in the peptide binding region, while sequence variation in pseudogenes would occur randomly (Hughes and Nei, 1989). Consistent with the amplification of expressed genes, the Wu-Kabat variability plots show that the most polymorphism occurs in residues that code for the peptide binding region in mammals, and corresponds well with the amplification of *Mhc* genes in *Barbus spp.* and the common carp (Dixon *et al.*, 1996). Furthermore, tests of selection indicated that positive or diversifying selection (dN > dS) is occurring in the peptide binding regions of the scaleless carp sequences, as seen in the expressed *Mhc* genes of many other taxa (Dixon *et al.*, 1996; Bernatchez and Landry, 2003).

As it is easier and more cost effective to type *Mhc* loci using genomic DNA rather than cDNA, future studies should involve a more thorough comparison of genomic and cDNA sequences to ensure that the genomic loci under investigation are expressed. This can be done by analyzing both cDNA and genomic DNA from the same individual, which was done on two samples in this study. This analysis confirmed that pseudogenes were not amplified, although more individuals should be screened in this fashion in order to confirm this result.

3.5.4 Information needed to fully characterize the *Mhc* in scaleless carp

As multiple sequences were found in the two individuals that were examined in detail (possibly representing different alleles), detailed sequencing analysis needs to be performed on many more clones. Many characterization studies utilize Rapid Amplification of cDNA Ends, or RACE-PCR, to obtain full-length transcripts of *Mhc* genes (Godwin *et al.*, 1997; Yu *et al.*, 2009). This procedure was attempted many times using the primers listed in this study; however it only resulted in non-specific product, or intriguingly, pyruvate dehydrogenase (D. O'Bryan, unpublished data).

Since *Mhc* in scaleless carp was found to be highly variable, and multiple sequences were detected in individuals for whom more than one clone was sequenced, a screening of many more samples should be performed. Further, multiple sequences isolated in this study differed by only one or two nucleotides. Additional clones from these individuals should be sequenced to confirm variation and possibly identify additional alleles. The identification of additional alleles could aid in the population genetics study in Lake Qinghai as it may be possible that certain alleles or loci coincide with fish sampled from different spawning sites (Beacham *et al.*, 2004).

4 GENERAL DISCUSSION

4.1 SPECIES, SUBSPECIES, AND ESUs

Although there are many species concepts, two were considered to be most relevant to this study – the Biological Species Concept and the Phylogenetic Species Concept (General Introduction). *Gymnocypris przewalskii* and *Gymnocypris eckloni* would not be considered separate species under the latter concept due to the lack of reciprocal monophyly; however, there is no physical chance, geographically speaking, that gene flow could ever occur between the two species, which would allow them to be defined as separate species under the Biological Species Concept.

Computer modelling has indicated that when two daughter populations separate from an ancestral gene pool, the phylogenetic relationships move chronologically through stages of polyphyly, paraphyly, and finally reciprocal monophyly (Neigel and Avise, 1986) (Figure 4.1). Using the assumptions of neutrality, the probabilities of daughter populations being polyphyletic, paraphyletic, or reciprocally monophyletic at a given time (measured in generations) are a function of the female effective population size. Under this model, reciprocal monophyly is most likely to occur when the number of generations is greater than four times the female effective population size (Avise, 2000). Using the high and low values of female effective population size for *G. p. przewalskii* (from Table 2.6: 71, 500 to 21, 050), reciprocal monophyly has a high probability to occur after 84, 200 to 286, 000 generations have elapsed, or 589, 400 to 2, 002, 000 years. Prior to this time, polyphyletic or paraphyletic daughter populations are more probable. Lake Qinghai was formed 130, 000 to 150, 000 years ago, equivalent to approximately 18, 000 to 21, 000 generations. Therefore, as stated in Chapter 2, it is not unexpected that *G. p.*

przewalskii and *G. eckloni* are currently paraphyletic at mtDNA alleles. Further, this lack of reciprocal monophyly at mtDNA alleles would also prevent the two species from even being considered separate ESUs under Moritz's (1994) definition. However, they occupy completely different habitats (saline Lake Qinghai vs. freshwater Yellow River) and are morphologically distinct. Indeed, the two most certainly should be considered both separate species and distinct ESUs.

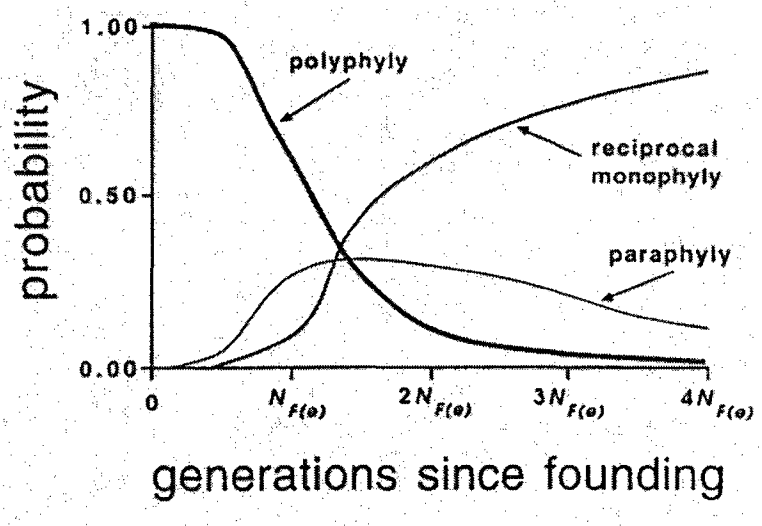


Figure 4.1: Probability curves for reciprocal monophyly, polyphyly, and paraphyly for two daughter populations following separation from an ancestral gene pool (after Neigl and Avise, 1986). Figure comes from Avise, 2000.

mtDNA and AFLP data are consistent with a subspecific classification of *G. p. ganzihonensis*, as AMOVAs using both markers indicated significant differences between the Ganzi fish and the Lake Qinghai fish (Chapter 2). However, additional studies are needed to determine if the subspecies should be considered a distinct ESU for conservation purposes. Generally speaking, the lack of reciprocal monophyly at mtDNA alleles would preclude the classification as a distinct ESU; however, this should not be expected for the reasons explained

above, and given the lack of reciprocal monophyly when comparing *G. p. przewalskii* to its most recent ancestor, *G. eckloni*.

Additionally, the AFLP and mtDNA evidence (high frequency of unique mtDNA haplotypes, clustering of some AFLP genotypes with *G. eckloni*) seem to indicate that the Ganzi River subspecies might have arisen not directly from the lake species after the Ganzi River was separated, but from its own evolutionary path, and would therefore represent "an important evolutionary legacy of a species" (Waples, 1991). Currently, the Ganzi River drains into a freshwater marsh that is no longer connected to the lake. It can, therefore, be speculated that the Ganzi River fish perhaps used to reside only within the river and marsh (which contains deep pools suitable for overwintering) and never actually utilized the lake. Interestingly, a recent study has found that a group of *G. eckloni* that reside in Lake Sunmcuo (which also used to be part of the Yellow River drainage) is genetically more similar (incomplete lineage sorting, but also a few shared haplotypes) to its sympatric subspecies, *G. e. scoliostomus*, than to *G. eckloni* of the Yellow River (Zhao *et al.*, 2009). The two subspecies are thought to have evolved in sympatry due to adaptations to different food sources.

Given the lack of reciprocal monophyly and the relatively recent ancestry of *G. przewalskii*, the principles of genetic and ecological exchangeability seem a better fit for determining ESU classifications in this case (Crandall *et al.*, 2000). AFLP and mtDNA data indicate a lack of genetic exchangeability; however, determining ecological exchangeability requires additional information. For instance, can the Ganzi River fish physiologically tolerate the saline waters of Lake Qinghai (Chapter 2)? As most cyprinids are generally considered stenohaline (de Boeck *et al.*, 2000; Luz *et al.*, 2008), it seems likely that they would. However, lethal sampling and subsequent tissue collection, often used in physiological studies (see Wood

et al., 2007), are not recommended. Therefore, any future physiological studies should be of minimal disruption to the fish, perhaps using blood collection only. Waples (1991) defined an ESU as being "substantially reproductively isolated" (without actually quantifying what is considered "substantial"). For fish populations, this may be timing or location of spawning. The subspecies was caught in only a single location within the river (which was quite difficult to locate); however, the range that the subspecies occurs within the river should be determined, and since the river itself has been so extensively altered due to irrigation and other anthropogenic factors, it should be mapped properly. The range could be determined simply by visual observation, or by other techniques such as the use of static nets and traps or electrofishing, which could determine the presence or absence of fish in a particular reach of the river (Lucas and Baras, 2000). Mark-recapture studies could also be conducted, which can estimate population size, mortality, and even growth rate, in addition to providing valuable information about the patterns of movement between the sites where fish were marked and then later recaptured. However, whether the Ganzi River fish can be considered an ESU or not, given its threatened status, the heavy anthropogenic impacts to the Ganzi River, and the general lack of information about this fish, *G. p. ganzihonensis* should *at least* be considered distinct Management Units (MUs), which requires monitoring and applying short-term management strategies (as recommended by Moritz, 1994).

AFLP data also indicate the presence of distinct stocks or MUs within Lake Qinghai (Chapter 2), although the detection of within-river structuring indicates that additional studies are needed, such as determining the timing and locations of spawning and calculating straying rates. Do they spawn annually, biannually, or go even longer between spawning events? Where are the important spawning areas, especially on the Buha River? Telemetry studies (for instance with

PIT tags) would be especially valuable in providing information about the migration patterns and range of the fish (Lucas and Baras, 2000). Tagging of juveniles in their natal rivers would also allow the determination of straying versus natal homing rates (Candy and Beacham, 2000).

4.2 USING *MHC* VARIATION TO DEVELOP A GENOTYPING SYSTEM

The ultimate goal of the *Mhc* portion of this thesis was to develop an additional genotyping tool that can be used to examine the spatial genetic relationships of the scaleless carp of Lake Qinghai. The results reported in Chapter 3 represent a first step in this analysis. Future studies will require the development of a primer pair that will target a single locus, which is necessary for *Mhc* genes to be fully utilized as markers to study scaleless carp populations. The common carp primers used in this study resulted in the amplification of more than one locus in certain individuals, in addition to non-specific product in most individuals. Additional work (not reported in this thesis) utilizing species-specific primers developed using scaleless carp exon 2 sequences described in Chapter 3 were not successful in amplifying a single locus, likely due to high level of sequence variation in this region (D. O'Bryan, unpublished data). Similarly, multiple attempts to sequence the full-length *Mhc* transcripts via RACE (Rapid Amplification of cDNA Ends) PCR mostly resulted in the amplification of non-specific product (D. O'Bryan, unpublished data).

Based on the results of this study, the *DAB3* locus group is not a good candidate for genotyping due to its prohibitive size, especially since only partial sequences of the first intron can be obtained from a single clone (two sequencing reactions, one in each direction, are necessary to get a full length sequence). However, sequence analysis of the first intron of the *DAB1* locus group revealed a series of short indels, which may be useful in discriminating *Mhc* genotypes. Future work could involve a more complete examination of the intronic region of

both locus groups, possibly designing primers that flank the intron in order to overcome sequence read length limitations (typically a maximum of 700 bp can be sequenced).

A study by an undergraduate student examined the utility of SSCP (Single Stranded Conformational Polymorphism) in identifying *Mhc* genotypes, specifically by attempting to amplify the exon 2 or the intron 1/exon 2 region of the *DAB1* locus group with additional species-specific primers (deRuiter, 2007). The intron1/exon 2 product was successfully amplified without multiple locus groups (*i.e.*, *DAB3* sequences) or non-specific products. SSCP results indicated substantial variation among samples, although the size of the product is larger than normally recommended for SSCP analysis and size variants were not easily distinguished. Optimization of the SSCP protocol, or an alternative method for screening (such as fragment analysis), may improve the resolution and allow the discrimination of size variants in scaleless carp samples. It may also be necessary to target sequence variation within intron 1 (excluding exon 2), or to design primers that span the intron 1-exon 2 boundary, before *Mhc* genes can be utilized as a polymorphic marker system. Additionally, segregation studies may also be required in order to accurately identify alleles and loci (Málaga-Trillo *et al.*, 1998; Murray *et al.*, 2000a; Murray *et al.*, 2000b), though such studies require genetically well-defined homozygous strains, produced via inbreeding by gynogenetic or androgenetic reproduction (van Erp *et al.*, 1996). Studies should also be performed to determine whether variation at *Mhc* loci and alleles are able to positively identify significant differences between scaleless carp spawning populations (Miller and Withler, 1997; Miller *et al.*, 2001).

Studies of *Mhc* sequences, macromutations (*i.e.,* indels), and allele frequencies can be used to study vertebrate phylogeny (Klein *et al.*, 1997); therefore, future works could involve the study of phylogeny and phylogeography of the 'snow carp' of the Tibetan Plateau. However,

Mhc alleles are known to persist in populations for very long times, and may even survive a speciation event, known as 'trans-species' evolution or polymorphism (Figueroa *et al.*, 2000). The use of *Mhc* in phylogenetic studies has therefore been criticized, as balancing selection and 'trans-species' evolution may cause discordance among gene trees and species trees (Edwards *et al.*, 2000). However, in combination with a neutral marker, like mtDNA which has been used in phylogenetic studies of cyprinids (Liu and Chen, 2003) and has been successfully sequenced in scaleless carp and related species (Chapter 2; Qi *et al.*, 2007; Zhao *et al.*, 2005), *Mhc* could be a powerful and effective tool, not only to construct a phylogeny of the snow carp, but to also examine the evolution of *Mhc* genes and explore the role of selective pressures in maintaining variation. A phylogeny of the east African haplochromine fishes has been constructed using mtDNA and *Mhc* genes (and also SINEs, or Short Interspersed Nuclear Elements, another neutral marker), with mtDNA used to determine common ancestry among fishes and *Mhc* used to explore the selective processes at work in historical fish populations (Sato *et al.*, 2003).

4.3 CONCLUSION

Scaleless carp are important for subsistence and economic development in a region that has a rapidly expanding human population. The long term survival and productivity of a harvested species rely not only on the conservation of numbers, but on the conservation of its genetic diversity, including its population structure (Nelson and Soule, 1987), so that MUs can be defined and the fishery managed accordingly. So far, the MUs have been determined by sampling fish during the spawning run. However, it may be possible that fish from different spawning populations co-mingle within the lake. If the lake fishery is to be reopened, the productivity and proportion of each spawning population should be determined (Flannery *et al.*, 2007). This can be accomplished using a genotyping tool that can identify each stock within the

lake. Multiple marker systems are often used to identify stocks in fish populations (Beacham *et al.*, 2004; Beacham *et al.*, 2005) and can offer a greater amount of precision in estimating population origins in mixed stock fisheries (Scribner *et al.*, 1998). Most studies used a combination of *Mhc* variation and microsatellites. However, in this case, there is the potential to develop a combination genotyping system using AFLPs and *Mhc* variation, even though AFLPs were initially used in this study to detect differences between fish samples from different spawning locations.

Traditionally, AFLPs have most commonly been used in plant, fungal, and bacterial studies and have generally been underused in animal studies, with microsatellites being the "gold standard" for detecting population level differences. While each microsatellite can have several alleles per locus, AFLPs have only two options per locus, present or absent. Therefore, each microsatellite locus has more power than a single AFLP locus. However, since a large number of AFLP loci can be screened in a single assay, AFLPs become a much more powerful tool. Recent studies indicate that AFLPs are at least as effective as microsatellites in the detection of differences between or among populations, with AFLPs being slightly better at detection of individual level differences and in assigning individuals to populations within admixed systems (Garoia *et al.*, 2007; Sonstebo *et al.*, 2007). Although population differences were detected between most spawning locations of scaleless carp, it may be necessary to use additional AFLP primer combinations (three primer combinations were subjectively chosen based on cost and time limitations) - six seems to be the "norm" for many studies (Sonstebo et al., 2007). When combined with *Mhc* genes, which will provide qualitative information on the adaptive variation present among stocks, AFLPs look to be very promising in developing an effective genotyping system to attempt to conserve this unique and threatened species

REFERENCES

April, J. & Turgeon, J. (2006). Phylogeography of the banded killifish (*Fundulus diaphanus*): glacial races and secondary contact. *Journal of Fish Biology* **69**, 212-228.

Avise, J. C. (2000). *Phylogeography: The History and Formation of Species*. Camridge, MA: Harvard University Press.

Banarescu, P. & Coad, B. (1991). Cyprinids of Eurasia. In *Cyprinid Fishes* (Winfield, I. & Nelson, J., eds.), pp. 127-155. London: Chapman and Hall.

Bartley, D., Bagley, M., Gall, G. & Bentley, B. (1994). Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. *Conservation Biology* **6**, 365-375.

Beacham, T. D., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen, K. H., Deng, L., Miller, K. M., Withler, R. E. & Varnavskaya, N. V. (2005). Estimation of stock composition and individual stock identification of sockeye salmon on a Pacific Rim basis using microsatellite and major histocompatibility complex variation. *Transactions of the American Fisheries Society* **134**, 1124-1146.

Beacham, T. D., Candy, J. R., Supernaulty, K. J., Ming, T., Deagle, B., Schulze, A., Tuck, D., Kaukinen, K. H., Irvine, J. R., Miller, K. M. & Withler, R. E. (2001). Evaluation and application of microsatellite and Major Histocompatibility Complex variation for stock identification of coho salmon in British Columbia. *Transactions of the American Fisheries Society* **130**, 1116-1149.

Beacham, T. D., Lapointe, M., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen, K. H., Deng, L., Miller, K. M. & Withler, R. E. (2004). Stock Identification of Fraser River Sockeye Salmon Using Microsatellites and Major Histocompatibility Complex Variation. *Transactions of the American Fisheries Society* **133**, 1117-1137.

Bensch, s. & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899-2914.

Bernatchez, L. & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16**, 363-377.

Bian, Q. T., Liu, J. Q., Luo, X. Q. & Xiao, J. L. (2000). Geotectonic setting, formation, and evolution of the Qinghai Lake. *Seismology and Geology* **22**, 20-26.

Brauner, C. J., Richards, J. G., Matey, V., Wood, C. M., Rogers, J., Murray, B. W., Chen, X. -Q., Du, J. -Z., & Wang, Y. -X. (2008). Salinity tolerance of the endangered Lake Qinghai scaleless carp, *Gymnocypris przewalskii*. *Comparative Biochemistry and Physiology, Part C* **148**, 447-448.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993). Three-dimensional structure of human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39.

Brunner, P. C., Douglas, M. R. & Bernatchez, L. (1998). Microsatellite and mitochondrial DNA assessment of population structure and stocking effects in Arctic charr *Salvelinus alpinus* (Teleostei: Salmonidae) from central Alpine lakes. *Molecular Ecology* 7, 209-223.

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

Chakroborty, R. & Weiss, K. M. (1991). Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. *American Journal of Human Genetics* **86**, 497-506.

Chen, D., Zhang, C., Lu, C., Chang, Y., & Chang, J. (2005). Amplified fragment length polymorphism analysis to identify the genetic structure of the *Gymnocypris przewalskii* (Kessler, 1876) population from the Qinghai Basin, China. *Journal of Applied Ichthyology* **21**, 178-183.

Clement, M., Posada, D. & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Cohen, S., Tirindelli, J., Gomez-Chiarri, M. & Nacci, D. (2006). Functional implications of Major Histocompatibility (MH) variation using estuarine fish populations. *Integrative and Comparative Biology* **46**, 1016-1029.

Cracraft, J. (1983). Species concepts and speciation analysis. In *Current Ornithology* (Johnston, R. F., ed.). New York: Plenum Press.

Crandall, K. A., Bininda-Edmonds, O. R. P., Mace, G. M. & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**, 290-295.

de Boeck, G., Vlaeminck, A., Van der Linden, A. & Blust, R. (2000). The energy metabolism of common carp (*Cyprinus carpio*) when exposed to salt stress: an increase in expenditure or effects of starvation? *Physiological and Biochemical Zoology* **73**, 102-111.

deRuiter, G. (2007). Investigation of Single Stranded Conformational Polymorphism and Melt Curve analysis for optimizing results for possible assessment of allelic intron 1 and exon 1 varation of MHC class IIB gene in Scaleless Carp (*Gymnocypris przewalskii*) of Lake Qinghai, China. *Undergraduate Thesis*. Prince George, BC: University of Northern British Columbia.

Dixon, B., Nagelkerke, L. A. J., Sibbing, F. A., Egberts, E. & Stet, R. J. M. (1996). Evolution of *MHC* class II β chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). *Immunogenetics* **44**, 419-431.

Dobzhansky, T. (1951). *Genetics and the Origin of the Species (3rd Edition)*. New York: Columbia University Press.

Edwards, S. V., Nusser, J. & Gasper, J. (2000). Characterization and evolution of Mhc genes from non-model organisms, with examples from birds. In *Molecular Methods in Ecology* (Baker, A. J., ed.), pp. 168-207. Cambridge: Blackwell Scientific.

Egger, B., Koblmüller, S., Sturmbauer, C. & Sefc, K. M. (2007). Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. *BMC Evolutionary Biology* 7, 137-151.

Elderkin, C. L., Perkins, E. J., Leberg, P. L., Klerks, P. L. & Lance, R. F. (2004). Amplified fragment length polymorphism (AFLP) analysis of the genetic structure of the zebra mussel, *Dreissena polymorpha*, in the Mississippi River. *Freshwater Biology* **49**, 1487-1494.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Ewen, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87-112.

Excoffier, L. G., Laval, G. & Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatiacs Online* **1**, 47-50.

Falush, D., Stephens, M. & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7, 574-578.

Figueroa, F., Mayer, W. E., Sültmann, H., O'hUigin, C., Tichy, H., Satta, Y., Takezaki, N., Takahata, N. & Klein, J. (2000). *Mhc* class II *B* evolution in East African cichlid fishes. *Immunogenetics* **51**, 556-575.

Flannery, B. G., Wenburg, J. K. & Gharrett, A. J. (2007). Evolution of mitochondrial DNA variation within and among Yukon River chum salmon populations. *Transactions of the American Fisheries Society* **136**, 232-240.

Frankham, R., Ballou, J. D. & Briscoe, D. A. (2002). *An Introduction to Conservation Genetics*. Cambridge, England: Cambridge University Press.

Garant, D., Dodson, J. J. & Bernatchez, L. (2000). Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology* **9**, 615-628.

Garoia, F., Guarniero, I., Grifoni, D., Marzola, S. & Tinti, F. (2007). Comparative analysis of AFLPs and SSRs efficiency in resolving population genetic structure of Mediterranean *Solea vulgaris*. *Molecular Ecology* **16**, 1377-1387.

Gharrett, A. J., Matala, A. P., Peterson, E. L., Gray, A. K. & Li, Z. (2005). Two genetically distinct forms of rougheye rockfish are different species. *Transactions of the American Fisheries Society* **134**, 242-260.

Giannasi, N., Thorpe, R. S. & Malhotra, A. (2001). The use of amplified fragment length polymorphism in determining species trees at fine taxonomic levels: analysis of a medically important snake, *Trimeresurus albolabris*. *Molecular Ecology* **10**, 419-426.

Godwin, U. B., Antao, A., Wilson, M. R., Chinchar, V. G., Miller, N. W., Clem, L. W. & McConnell, T. J. (1997). MHC class II B genes in the channel catfish (*Ictalurus punctatus*). *Developmental and Comparative Immunology* **21**, 13-23.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704.

Haig, S. M., Beever, E. A., Chambers, S. M., Draheim, H. M., Dugger, B. D., Dunham, S., Elliott-Smith, E., Fontaine, J. B., Kesler, D. C., Knaus, B. J., Lopes, I. A., Loschl, P., Mullins, T. D. & Sheffield, L. M. (2006). Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. *Conservation Biology* **20**, 1584-1594.

Hawkins, J. S., Pleasants, J. & Wendel, J. F. (2005). Identification of AFLP markers that discriminate between cultivated cotton and the Hawaiian islands endemic, *Gossypium tomentosum* Nuttall ex Seeman. *Genetic Resources and Crop Evolution* **52**, 1069-1078.

Hayashi, E., Chi, H. C., Boyer, S. K. & Still, D. W. (2005). Amplified Fragment Lenght Polymorphism Protocol for Plant Science on CEQ Series Genetic Analysis System. Beckman Coulter, Inc.

He, D., Chen, Y. F., Chen, Y. Y. & Chen, Z. (2004). Molecular phylogeny of the specialized schizothoracine fishes (Teleostei: Cyprinidae),with their implications for the uplift of the Qinghai-Tibetan Plateau. *Chinese Science Bulletin* **49**, 39-48.

Hedrick, P.W. (2005). Genetics of Populations, Third Edition. Sudbury, MA: Jones and Bartlett Publishers.

Hey, J. (2001). The mind of the species problem. *Trends in Ecology and Evolution* **16**, 326-329.

Hughes, A. L. & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Science* **86**, 958-962.

Ishikawa, S., Tsukamoto, K. & Nishida, M. (2004). Genetic evidence for multiple geographic populations of the giant mottled eel *Anguilla marmorata* in the Pacific and Indian oceans. *Ichthyological Research* **51**, 343-353.

Klein, J. (1987). Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* **19**, 155-162.

Klein, J., Figueroa, F., Klein, D., Sato, A. & O'Huigin, C. (1997). Major histocompatibility complex genes in the study of fish phylogeny. In *Molecular Systematics of Fishes* (Kocher, T. D. & Stepier, C. A., eds.). San Diego, CA: Academic Press.

Kruiswijk, C. P., Hermsen, T., Fujiki, K., Dixon, B., Savelkoul, H. F. B. & Stet, R. J. M. (2004). Analysis of genomic and expressed major histocompatibility class Ia and class II genes in a hexaploid Lake Tana African 'large' barb individual (*Barbus intermedius*). *Immunogenetics* **55**, 770-781.

Kruiswijk, C. P., Hermsen, T., van Heerwaarden, J., Dixon, B., Savelkoul, H. F. B. & Stet, R. J. M. (2005). Major histocompatibility genes in the Lake Tana African large barb species flock: evidence for complete partitioning of class II *B*, but not class I, genes among different species. *Immunogenetics* **56**, 894-908.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics* **5**, 150-163.

Li, J. J. & Fang, X. M. (1998). Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin* **44**, 2117-2125.

Liu, H. & Chen, Y. (2003). Phylogeny of the East Asian cyprinids inferred from sequences of the mitochondrial DNA control region. *Canadian Journal of Zoology* **81**, 1938-1946.

Lucas, M. C. & Baras, E. (2000). Methods for studying spatial behaviour of freshwater fishes in the natural environment. *Fish and Fisheries* **1**, 283-316.

Luz, R. K., Martinez-Alvarez, R. M., De Pedro, N. & Delgado, M. J. (2008). Growth, food intake regulation and metabolic adaptations in goldfish (*Carassius auratus*) exposed to different salinities. *Aquaculture* **276**, 171-178.

Málaga-Trillo, E., Zaleska-Rutczynska, Z., McAndrew, B., Vincek, V., Figueroa, F., Sültmann, H. & Klein, J. (1998). Linkage relationships and haplotype polymorphism among cichlid *Mhc* class II *B* loci. *Genetics* **149**, 1527-1537.

Matey, V., Richards, J. G., Wang, Y., Wood, C. M., Rogers, J., Semple, J. W., Murray, B. W., Chen, X. Q., Du, J. & Brauner, C. J. (2008). The effect of hypoxia on gill morphology and ionoregulatory status in the endangered Lake Qinghai scaleless carp, *Gymnocypris przewalskii*. *Journal of Experimental Biology* **211**, 1063-1074.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. & Ashlock, P. A. (1991). Principles of Systematic Zoology. New York: McGraw-Hill.

Meng, P., Wang, W.-J., Kong, J., Li, X.-J. & Shi, J.-Q. (2007). Isozyme variation of Lake Qinghia naked carp *Gymnocypris przewalskii* (Kessler) from five rivers. *Acta Zoologica Sinica* **53**, 892-898.

Mickett, K., Morton, C., Feng, J., Li, P., Simmons, M., Cao, D., Dunham, R. A. & Liu, Z. (2003). Assessing genetic diversity of domestic populations of channel catfish (*Ictalurus punctatus*) in Alabama using AFLP markers. *Aquaculture* **228**, 91-105.

Miller, K. M., Beacham, T. D. & Withler, R. E. (2001). Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111**, 237-257.

Miller, K. M. & Withler, R. E. (1996). Sequence analysis of a polymorphic *Mhc* class II gene in Pacific salmon. *Immunogenetics* **43**, 337-351.

Miller, K. M. & Withler, R. E. (1997). Mhc diversity in Pacific salmon: Population structere and trans-species allelism. *Hereditas* **127**, 83-95.

Miller, K. M., Withler, R. E. & Beacham, T. D. (1997). Molecular evolution at MHC genes in two populations of Chinook salmon (*Onchorhynchus tshawytscha*). *Molecular Ecology* **6**, 937-954.

Miller, M. P. (1997). Tools for population genetic analyses (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Milner, A. M., Knudsen, E. E., Soiseth, C., Robertson, A. L., Schell, D., Phillips, I. T. & Magnusson, K. (2000). Colonization and development of stream communities across a 200-year gradient in Glacier Bay National Park, Alaska, U.S.A. *Canadian Journal of Fisheries and Aquatic Science* **57**, 2319-2335.

Mock, K. E. & Miller, M. P. (2005). Patterns of molecular diversity in naturally occurring and refugial populations of the least chub. *Transactions of the American Fisheries Society* **134**, 267-278.

Moritz, C. (1994). Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution* **9**, 373-375.

Mueller, U. G. & Wolfenbarger, L. L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**, 389-394.

Murray, B. W., Nilsson, P., Zaleska-Rutczynska, Z., Sültmann, H. & Klein, J. (2000a). Major histocompatibility complex class I *A* genes in the cichlid fish *Oreochromis* niloticus. *Marine Biotechnology* **2**, 437-448.

Murray, B. W., Shintani, S., Sültmann, H. & Klein, J. (2000b). Major histocompatibility complex class II *A* genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. *Immunogenetics* **51**, 576-586.

Nei, M. (1972). Genetic distance between populations. *American Naturalist* **106**, 283-292.

Nei, M. & Hughes, A. L. (1991). Polymorphism and the evolution of the major histocompatibility complex loci in mammals. In *Evolution at the Molecular Level* (Selander, R. K., Clark, A. G. & Whittam, T. S., eds.), pp. 222-247. Sunderland, MA: Sinauer Associates, Inc.

Neigel, J. E. & Avise, J. C. (1986). Phylogenetic relationships of mitochrondrial DNA under various demographic models of speciation. In *Evolutionary Processes and Theory* (Nevo, E. & Karlin, S., eds.), pp. 515-534. New York: Academic Press.

Nelson, K. & Soule, M. (1987). Genetical conservation of exploited fishes. In *Population genetics and fishery management* (Ryman, N. & Utter, F., eds.), pp. 345-368. Seattle: University of Washington Press.

Pascual, M. A., Quinn, T. P. & Fuss, H. (1995). Factors affecting the natal homing of fall chinook salmon from the Columbia River hatcheries. *Transactions of the American Fisheries Society* **124**, 308-320.

Peakall, R. & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Perry, S. F. & Laurent, P. (1989). Adaptational responses of rainbow trout to lowered external NaCl: contribution of the branchial chloride cell. *Journal of Experimental Biology* **147**, 147-168.

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Postlethwaite, E. K. & McDonald, D. G. (1995). Mechanisms of Na^+ and Cl^- regulation in freshwater-adapted rainbow trout (*Oncorhynchus mykiss*) during exercise and stress. *Journal of Experimental Biology* **198**, 295-304.

Qi, D., Guo, S., Tang, J., Zhao, X. & Liu, J. (2007). Mitochondrial DNA phylogeny of two morphologically enigmatic fishes in the subfamily Schizothoracinae (Teleostei: Cyprinidae) in the Qinghai-Tibetan plateau. *Journal of Fish Biology* **70**, 60-74.

Qi, D., Li, T., Zhao, X., Guo, S. & Li, J. (2006). Mitochondrial cytochrome b sequence variation and phylogenetics of the highly specialized Schizothoracine fishes (Teleosteii: Cyprinidae) in the Qinghai-Tibet Plateau. *Biochemical Genetics* **44**, 270-285.

Quinn, T. P. & Dittman, A. H. (1990). Pacific salmon migrations and homing: mechanism and adaptive significance. *Trends in Ecology and Evolution* **5**, 174-177.

Rakus, K. Ł., Wiegertjes, G. F., Stet, R. J. M., Savelkoul, H. F. J., Pilarczyk, A. & Irnazarow, I. (2003). Polymorphism of major histocompatibility complex class II *B* genes in different lines of the common carp (*Cyprinus carpio*). *Aquatic Living Resources* **16**, 432-437.

Rodrigues, P. N. S., Dixon, B., Roelofs, J., Rombout, J. H. M. W., Egberts, E., Pohajdak, B. & Stet, R. J. M. (1998a). Expression and temperature-dependent regulation of the $Beta_2$-Microglobulin (*Cyca-B2m*) gene in a cold-blooded vertebrate, the common carp (*Cyprinus carpio* L.). *Developmental Immunology* **5**, 263-275.

Rodrigues, P. N. S., Hermsen, T. T., Rombout, J. H. M. W., Egberts, E. & Stet, R. J. M. (1995). Detection of MHC Class II Transcripts in lymphoid tissues of the common carp (*Cyprinus carpio* L.). *Developmental and Comparative Immunology* **19**, 483-496.

Rodrigues, P. N. S., Hermsen, T. T., van Maanen, A., Taverne-Thiele, A. J., Rombout, J. H. M. W., Dixon, B. & Stet, R. J. M. (1998b). Expression of *MhcCyca* class I and class II molecules in the early life history of the common carp (*Cyprinus carpio* L.). *Developmental and Comparative Immunology* **22**, 493-506.

Ryder, O. A. (1986). Species conservation and sytematics: the dilemma of subspecies. *Trends in Ecology and Evolution* **1**, 9-10.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual (Third Edition)*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sato, A., Takezaki, N., Tichy, H., Figueroa, F., Mayer, W. E. & Klein, J. (2003). Origin and speciation of haplochromine fishes in East African crater lakes investigated by the analysis of their mtDNA, *Mhc* genes, and SINEs. *Molecular Biology and Evolution* **20**, 1448-1462.

Scribner, K. T., Crane, P. A., Spearman, W. J. & Seeb, L. W. (1998). DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). *Canadian Journal of Fisheries and Aquatic Science* **55**, 1748-1758.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. J. M. & Bills, R. (2003). Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *Proceedings of the Royal Society of London B* **270**, 129-137.

Seutin, G., White, B. N. & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**, 82-90.

Sonstebo, J. H., Borgstrom, R. & Heun, M. (2007). A comparison of AFLPs and microsatellites to identify the population structure of brown trout (*Salmo trutta* L.) populations from Hardangervidda, Norway. *Molecular Ecology* **16**, 1427-1438.

Staines, N., Brostoff, J. & James, K., eds. (1993). *Introducing Immunology, Second Edition.* St. Louis: Mosby-Yearbook Europe Limited.

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

Stewart, I. J., Quinn, T. P. & Bentzen, P. (2003). Evidence for fine-scale natal homing among island beach spawning sockeye salmon, *Oncorhynchus nerka. Environmental Biology of Fishes* **67**, 77-85.

Sullivan, J. P., Lavoué, S., Arnegard, M. E. & Hopkins, C. D. (2004). AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* **58**, 825-841.

Tabata, K. & Taniguchi, N. (2000). Differences between *Pagrus major* and *Pagrus auratus* through mainly mtDNA control region analysis. *Fisheries Science* **66**, 9-18.

Tajima, F. (1983). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rae varies among sites. *Genetics* **143**, 1457-1465.

Taylor, E. B., Rendenbach, Z., Costello, A. B., Pollard, S. & Pacas, C. J. (2001). Nested analysis of genetic diversity in northwestern North American char, Dolly Varden (*Salvelinus malma*) and bull trout (*Salvelinus confluentus*). *Canadian Journal of Fisheries and Aquatic Science* **58**, 406-420.

van Erp, S. H. M., Egberts, E. & Stet, R. J. M. (1996). Characterization of class II *A* and *B* genes in a gynogenetic carp clone. *Immunogenetics* **44**, 192-202.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Walker, K. F., Dunn, I. G., Edwards, D., Petr, T. & Yang, H. Z. (1996). A fishery in a changing lake environment: the naked carp Gymnocephalus przewalskii (Kessler) (Cyprinidae: Schizothoracinae) in Qinghai Hu, China. *International Journal of Salt Lake Research* **4**, 169-222.

Walker, K. F. & Yang, H. Z. (1999). Fish and fisheries of western China. In *Fish and Fisheries at Higher Altitudes: Asia*, p. 304. Rome.

Wang, Y. S., Gonzalez, R. J., Patrick, M. L., Grosell, M., Zhang, C., Feng, Q., Du, J., Walsh, P. J. & Wood, C. M. (2003). Unusual physiology of scale-less carp, *Gymnocypris przewalskii*, in Lake Qinghai: a high altitude alkaline saline lake. *Comparative Biochemistry and Physiology Part A* **134**, 409-421.

Waples, R. S. (1991). Pacific Salmon, *Oncorhynchus* spp. and the definition of 'species' under the endangered species act. *Marine Fisheries Reviews* **53**, 11-22.

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7, 256-276.

Wehrhahn, C. F. & Powell, R. (1987). Electrophoretic variation, regional differences, and gene flow in the Coho Salmon (*Oncorhynchus kisutch*) of southern British Columbia. *Canadian Journal of Fisheries and Aquatic Science* **44**, 822-831.

Whitehead, A., Anderson, S. L., Kuivila, K. M., Roach, J. L. & May, B. M. (2003). Genetic variation among interconnected populations of *Catostomus occidentalis*: implications for distinguishing impacts of contamints from biogeographical structuring. *Molecular Ecology* **12**, 2817-2833.

Wood, C. M., Du, J., Rogers, J., Brauner, C. J., Richards, J. G., Semples, J. W., Murray, B. W., Chen, X. Q. & Wang, Y. (2007). Przewalski's Naked Carp (*Gymnocypris przewalskii*): An endangered species taking a metabolic holiday in Lake Qinghai, China. *Physiological and Biochemical Zoology* **80**, 59-77.

Wu, T. T. & Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *Journal of Experimental Medicine* **132**, 211-250.

Wu, Y. F. (1984). Systematics and taxonomy of the Schizothoracine fishes in China. *Acta Zoologica Plateau Sinica* **3**, 119-139.

Xie, Z., Du, J., Chen, X., Wang, Y. & Murray, B. (2006). The Significance of Mitochondria Control Region (D-Loop) in Intraspecific Genetic Differentiation of Fish. *Hereditas (Beijing)* **28**, 362-368.

Yang, J., Qi, H., Shi, J., Chen, D., Huang, F. & Tang, H. (2005). Hydrochemistry properties and water quality analysis of Qinghai hu lake. *Freshwater Fisheries* **35**, 28-32.

Young, W. P., Ostberg, C. O., Keim, P. & Thorgaard, G. H. (2001). Genetic characterization of hybridization and introgression between anadromous rainbow trout (*Onchorhyncus mykiss irideus*) and coastal cuthroat trout (*O. clarki clarki*). *Molecular Ecology* **10**, 921-930.

Yu, P. & Chen, Y. (1998). *China Red Data Book of Endangered Animals-Pisces*. Beijing: Science Press.

Yu, S. H., Ao, J. Q. & Chen, X. H. (2009). Molecular characterization and expression analysis of MHC class A and B genes in large yellow croaker (*Pseudosciaena crocea*). *Molecular biology Reports* **37**, 1295-1307.

Yuan, Z.K.X. (2005). Qing Zang gao yuan di tu ji (Mandarin Chinese Edition). ISBN: 7-03-003397-3

Zhan, T. (1994). Annex 2: Priority Species in China. In *China Biodiversity Conservation Action Plan* (Maxey, C. & Lutz, J., eds.). Beijing.

Zhang, L., Jiang, P., Gang, H., Shi, J. -Q., & Sun, X. (2009). Development and characterization of microsatellite markers to study population genetic diversity of Przewalski's naked carp *Gymnocypris przewalskii* in China. *Journal of Fish Biology* **75**, 2532-2536.

Zhao, K., Duan, Z. Y., Peng, Z. G., Guo, S. C., Li, J. B., He, S. P., & Zhao, X. Q. (2009). The yongest split in sympatric schizothoracine fish (Cyprinidae) is shaped by ecological adaptations in a Tibetan plateau glacier lake. *Molecular Ecology* **18**, 3616-3628.

Zhao, K., Duan, Z. Y., Yang, G. S., Peng, Z. G., He, S. P., & Chen, Y. Y. (2007). Origin of *Gymnocypris przewalskii* and phylogenetic history of *Gymnocypris eckloni. Progress in Natural Science* **17**, 520-528.

Zhao, K., Li, J., Yang, G., Duan, Z., He, S. & Chen, Y. (2005). Molecular phylogenetics of *Gymnocypris* (Teleostei: Cyprinidae) in Lake Qinghai and adjacent drainages. *Chinese Science Bulletin* **50**, 1325-1333.

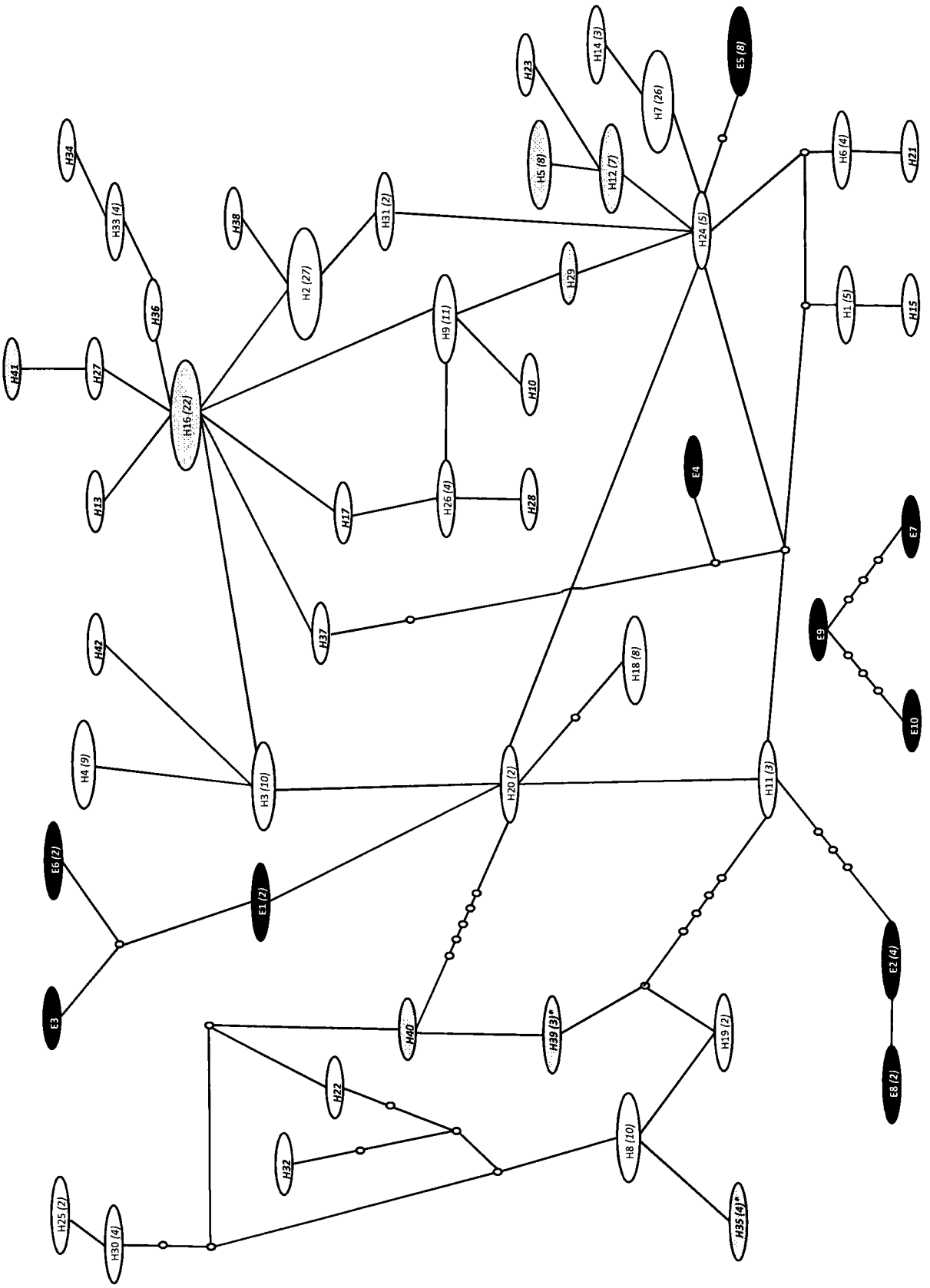
Zheng, M. (1997). An introduction to saline lakes on the Qinghai-Tibet plateau. In *Monographiae Biologicae* (Dumont, H. J. a. W., M.J.A., ed.), pp. 1-289. Beijing: Scientific Publisher.

Zhu, S. & Wu, F. (1975). Study of fish fauna in Qinghai Lake. In *Fish Fauna and Biology of Naked Carp in Qinghai Lake*, pp. 9-26. Beijing: Scientific Publisher.

Zouros, E. (1979). Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics* **92**, 623-646.

APPENDIX 1 – mtDNA PARSIMONY NETWORK

Figure A1.1: Parsimony network of *G. przewalskii* and *G. eckloni* mtDNA control region haplotypes. Grey shading indicates haplotypes found in the Ganzi River (* indicates haplotypes unique to the Ganzi River found in more than one individual), black shading indicates *G. eckloni* haplotypes, numbers in brackets indicate the number of times each haplotype is found in sequenced individuals, circles indicate mutational changes between haplotypes. Haplotypes thatdiffer by more than 7 mutational changes are not connected to the network.



APPENDIX 2 – DETAILED AFLP METHODOLOGY

Genomic DNA extractions

Fin clips were first rinsed of residual salts from the DMSO solution with Milli-Q water. Excess water was removed by blotting with a Kim Wipe. Fin clips were then ground into a paste in a porcelain mortar and pestle (cleaned first with hot soapy water, rinsed with distilled water followed by 70% ethanol). 500 µL of Queen's Lysis Buffer was added directly to the mortar and pestle. The tissue/buffer solution was decanted into a 1.5 mL microcentrifuge tube. Proteinase K (5 µL of 20 mg/mL) was added, followed by brief vortexing and incubating the samples at 50°C for one hour, after which a second spike (5 µL) of 20 mg/mL Proteinase K was added. To facilitate additional protein degradation, the solutions were incubated at 37°C overnight on a nutator.

Standard phenol/chloroform isolations were used for DNA extraction. After briefly spinning down the tissue lysis solutions, 600 µL of a phenol/chloroform/isoamyl alcohol solution (25:24:1) were added. The tubes were gently mixed by gentle inversion for at least three to four minutes. Tubes were centrifuged at 13,000 rpm for four minutes. The top organic phase was carefully removed without disturbing the interface layer. 500 µL of the phenol/chloroform/isoamyl alcohol solution were added to each tube, followed by gentle mixing and spinning (as described previously). The top phase was again removed, after which 450 µL of chloroform were added, followed by gentle mixing, centrifuging and removal of the top layer.

Two volumes of cold 95% ethanol and a one tenth volume of 3M sodium acetate were added to the solutions to precipitate the DNA. After gentle mixing, tubes were centrifuged at 13,000 rpm for thirty minutes to pellet the DNA. The ethanol was carefully drawn off, leaving

behind the DNA pellet. The pellet was rinsed twice with 500 µL of cold 70% ethanol. Each rinse was followed by a 4 minute centrifugation at 13,000 rpm. After the final rinse, the residual ethanol was drawn off and the pellet was allowed to dry. The DNA pellet was re-suspended in 100 µL of TE buffer (1M, pH 8.0) and frozen at -30°C.

DNA quality and quantity

The quality and quantity of the DNA was checked using two methods – agarose gel electrophoresis and spectrophotometry. In the first method, 8 µL of DNA combined with 2 µL of bromophenol blue gel loading dye was loaded onto a 0.7% agarose gel stained with ethidium bromide. The gel was run at 100 volts for approximately one hour. *Hae* III (Invitrogen) was used as a size and mass standard. High quality DNA was represented by a high molecular weight band versus the "smear" that results from sheared, low quality DNA. DNA quantity was also estimated by comparing the band intensities of the DNA samples to the intensities of the fragments in the mass standard. In the second method, 2 µL of DNA were loaded onto the NanoDrop spectrophotometer. The A_{260}/A_{280} ratio (approximately 1.8) and the graphics profile (clear and clean) were used to determine the presence of good quality DNA. Quantity was automatically calculated using the A260 value. As long as the NanoDrop quantity approximately matched the concentration estimated from the gel, the NanoDrop concentration was used to calculate DNA working stocks of 50 ng/uL.

AFLP Methods

The AFLP protocol was based on Vos (1995) and Hawkins *et al.* (2005) with modifications as recommended by Hayashi and colleagues (2005). See Table 2.1 for all adapter and primer sequences.

Polymorphism Screening

Sixteen selective primer combinations were chosen based on the primer trials of Garbalena (2005) who performed an AFLP study on common carp (*Cyprinus carpio*). Five genomic DNA samples each from the Heima, Buha, Shaliu, Yellow, and Ganzi Rivers were randomly chosen and taken through the entire AFLP method below. The best three primer combinations were chosen for the remainder of the samples based on observed levels of polymorphism (Table A2.1).

Table A2.1: Selective primer test matrix (X indicates primer combination used, **X** (in bold) indicates a primer combination used in data collection).

	E-AAG	E-ACT	E-AGC	E-ACG
M-CAA	X	X		**X**
M-CTG	X		X	X
M-CAC	**X**			X
M-CTT	X		X	X
M-CAA	X			
M-CAT		X		
M-CAG		X	X	**X**

The following primer combinations were used and labelled with the appropriate Beckman dyes (from Integrated DNATechnologies): D2-E-ACG/M-CAA, E-ACG/D3-M-CAG, and E-AAG/D4-M-CAC.

Digestion

Two hundred ng of genomic DNA were digested with 10 units each of *Mse*I and *Eco*RI, 1X *Eco*RI Restriction Buffer, and 2 μg of Bovine Serum Albumim (BSA) (New England

Biolabs). The solutions were incubated at 37°C on a PTC-100 Thermal Cycler for 3 hours, followed by 20 minutes at 65°C to inactivate the enzymes.

Ligation

Adapters for each restriction enzyme (50 µM final concentration) were prepared by mixing equal parts of forward and reverse adapters (100 µM) and heating at 95°C for 5 minutes, followed by slow cooling in a Styrofoam box at room temperature. The following master mix (20 µL) was added to each digestion: 3 units (NE Units) of T4 DNA Ligase, 1X T4 Ligase Buffer (New England Biolabs), and 1.5 µL each of 50 µM adapters (*Mse*I and *EcoR*I). Solutions were incubated at 16°C overnight (for at least 16 hours) in a PTC-100 Thermal Cycler, followed by 10 minutes at 65°C to inactivate the T4 ligase. Ligation solutions were diluted 1:5 with nuclease-free water.

Pre-amplification

Five µL of the diluted digestion-ligation product was added to 20 µL of the following master mix: 1X PCR Buffer, 1.5mM MgCl2, 200 µM of each dNTP, 0.8uM each of the *EcoR*I-A and *Mse*I-C preamplification primers, and 1.25 units of *Taq* polymerase (Invitrogen). Samples were run on a Dyad Thermal Cycler with the following conditions: 72°C for 2 minutes followed by 20 cycles of 95°C for 30 seconds, 56°C for 1 minute, and 72°C for 1 minute; 60°C for 30 minutes, and held at 4°C. The pre-amplification product was diluted 1:10 in nuclease-free water (IDT).

Selective amplification

Five µL of the diluted pre-amplification product was added to 15 µL of the following master mix: 1X PCR Buffer, 1.5mM MgCl2, 200 µM of each dNTP, 0.4 µM each of the EcoRI and MseI selective primers (see Table), and 1 unit of Platinum Taq (Invitrogen). Samples were run on a Dyad thermal cycler with the following conditions: an initial activation of 94°C for 30 seconds, followed by 12 cycles of 94°C for 30 seconds, 65°C (decreased by 0.7°C per cycle) for 30 seconds, and 72°C for 1 minute. The preceding 13 cycles were followed by 32 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute. The program was finished with a final extension phase of 72°C for 10 minutes followed by holding at 4°C. Amplification products were stored at -30°C until loaded on the Beckman CEQ8000.

Quality Assurance/Quality Control

Prior to digestion, all genomic DNA samples were checked for quality and quantity using gel electrophoresis (0.7% agarose gel) and spectrophotometry (Nanodrop). Only high quality DNA was used as evidenced by a high molecular weight band on the agarose gel rather that a "smear" of low quality sheared DNA.

Generally AFLPs are considered to be a non-specific technique; therefore, it was extremely important to ensure that all reagents are free of contaminants. Only fresh reagents dedicated for AFLP use were utilized. All primers, adapters, and solutions were reconstituted and diluted using sterile nuclease-free water (IDT). All steps of the process were performed in a sterile PCR hood, using dedicated pipettors and freshly autoclaved barrier tips and tubes.

Genomic DNA samples were processed in "batches" of 43. To reduce bias, samples from one river were always run in the same batch as samples from at least one other river. A

digestion-ligation blank (nuclease-free water added to the digestion master mix) was processed with each batch to check for contamination in any of the reagents. This blank was carried through each AFLP step. Repeatability of the entire AFLP methodology was checked by randomly choosing one sample to run in triplicate in the digestion-ligation procedure. Repeatability was further assured by running each digestion-ligation product (including the blank and the samples run in triplicate) in duplicate through the pre-amplification and selective amplification procedures.

Fragment Sizing

AFLP fragments were sized using the Beckman CEQ8000. Each of the three selective amplification products were combined (0.5 μL D2, 1μL D3, and 0.5 μL D4) and added to 37.5 μL Sample Loading Solution (deionized formamide) and 0.5 μL 600 bp size standard. Samples were run using the Frag-2 separation method.

Fragments were viewed in the CEQ8000 Fragment Analysis Module. Fragment sizes were generated using a 600 bp size standard, model set to quartic, and the dye mobility calibration set to PA ver.1.

Each sample's replicate was scored. If a fragment did not occur in each replicate, it was removed from the analysis. Sample pairs with several fragments which were not in agreement were re-analyzed. A reliable locus list was generated by examining all sample duplicates and determining which fragments were repeatable and which were not. Reliable fragments had strong intensities, were clear and crisp, and were greater than 80 bp in size. If a fragment was not considered reliable or was consistently unrepeatable, it was dropped from analysis. Any

fragments which were observed in blank samples were also removed from analysis. This approach was done to ensure a conservative, but very reliable data set.

Fragments were binned in the AFLP analysis module of the Beckman Coulter software using the following parameters: bin of 1, no Y threshold was set, since all samples were manually scored anyways, fully populated bins excluded. A binary matrix was generated and exported as an Excel spreadsheet.

APPENDIX 3 – *MHC* SEQUENCES

Intron 1

```
Gypr-DAB1-16(g)   GTAAATCAAATCAAGTGTAATACAAAATTA-ATATATATATATATTTACAATAACATTATTCATGATTGATTACTATTGTTAGTAGAAGTAGAAGTAAA
Gypr-DAB1-14(g)   ..............................-...--------.....................................A....A......------...
Gypr-DAB1-15(g)   ..............................-...--------.....................................A....A...G..------...
Gypr-DAB1-13(g)   ..............................-...--------.....................................A....A......------...
Gypr-DAB1-11(g)   ....C......GT.A.AC..A.TGT.TAG.T...--------...........A.A.....................................
Gypr-DAB1-08(g)   ..............C.............-...--------.................A.....................................
Gypr-DAB1-07(g)   ............................G-...G.....--.....................................................
Gypr-DAB1-05(g)   ..............................-...--------.....................................................
Gypr-DAB1-06(g)   ..............................-...--------.....................................................
Gypr-DAB1-02d(g)  ............................A.ATA---------....C.......G.....................................------...
Gypr-DAB1-02      ............................ATA---------....C.......G.....................................------...
Gypr-DAB1-01      ..........AT.TA.A..TATAT.T..ATATA---------.....................................................

Babo-DAB1         ......AC...T..............G.-...G.....--.....................................................T......
Bain-DAB1-06      ......AC..AT..............A.-...G..T..--.....................................T....A..........---......
Bain-DAB1         ......AC...T..............A.-...G..T..--.....................................G..................C.T......
Cyca-DAB1         ......A....T...C.....GT.....A.-G..G.....--..C.G...G..................A....T..................T......
Cyca-DAB2         ......A....T...C.....GT.......-...G.....--....G...G.......T........GA....T..................T......


Gypr-DAB1-16(g)   CAGGACAGGTCACTAAAACAGCATCCTA----TGTATCTAA-GCCCTTCTGTTTTTCTAGAGGTGACATGACATCATCTGAAATAACT--AATGCTGTT
Gypr-DAB1-14(g)   .........G..........A..G..C.----.........-.T.....................................................--....A..A.
Gypr-DAB1-15(g)   .........G..........A..A..C.----.........-.T.....................................................--....A..A.
Gypr-DAB1-13(g)   .........G..........A..A..C.----.........-.T.....................................................--....A..A.
Gypr-DAB1-11(g)   ..................T.A..G..C.----.........-.....................................................--.........
Gypr-DAB1-08(g)   ..................A..A..C.----.........-.....................--.....................................--.........
Gypr-DAB1-07(g)   .........G..........A..A..C.----......G..-.....................................---.................--.........
Gypr-DAB1-05(g)   ..................A..A..C.----....AT...-.....................................................--.........
Gypr-DAB1-06(g)   .........G..............C.----.........-.....................................................--.........
Gypr-DAB1-02d(g)  .........G..........A..A..C.----.........-.....................................................--.........
Gypr-DAB1-02      .........G..........A..A..C.----.........-.....................................................--.........
Gypr-DAB1-01      .........G..........A..A..C.----.........-.....................................................--.........

Babo-DAB1         ..................A..G..A.CCTT.........GC.....G.....................................--....A....
Bain-DAB1-06      ..................A..A..A.TCTC...G...G.-C...........C........G...G....G....G..........--G...A....
Bain-DAB1         ..................A..G..A.TCTC...G...G.-...........C.............G....G..............--....A....
Cyca-DAB1         ..A.....T.....---.T.T.A..A.CTAA........G-.T..............T..T........A.........C......GT.......C.
Cyca-DAB2         ..A.....T.....---.T.T.A..A.CTAA........G-.T..............T..T........A.........C......GT.......C.
```

Exon 2

```
Gypr-DAB1-16(g)    TTATCATTACTTT-CAGCTGATGGATACTACAATTCTCGCTGGGCTAAATGCATCCACAGCTCCCGTGATCTCACTGACATCGTGTTCGTTGATAACTA
Gypr-DAB1-14(g)    ...A.........-.................C.....AC....AT..G....T.....................G...T.........A.C........
Gypr-DAB1-15(g)    ...A.........-.................C.....AC....AT..G....T.....................G...T.........A.C........
Gypr-DAB1-13(g)    ...A.........-.................C.....AC....AT..G....T.....................G...T.........A.C........
Gypr-DAB1-12(c)                  ...............C.....AC....AT..G....T.....................G...T.........A.C........
Gypr-DAB1-11(g)    ...A.........-.................C.....AC....AG..G.........................G......G....A.A.C........
Gypr-DAB1-09(c)                  ...............C...........A...G.......T.................................A.C........
Gypr-DAB1-10(c)                  ...............T.....TT....AG..G.................................G....A.A.C........
Gypr-DAB1-08(g)    ...A.........-.................T...........G..G.................................G....A.A..........
Gypr-DAB1-07(g)    .............-.....................A.GTT...AA.............................G......G..................
Gypr-DAB1-05(g)    ...A.........-.................C...........A...G.................................G...A.GA..........
Gypr-DAB1-06(g)    ...A.........T.................G...ATTTA..AGA.G.........................G......G..........C.....C
Gypr-DAB1-02d(g)   .-.A.........-.................G....TAT...A...G.....................................................
Gypr-DAB1-02       .-.A.........-.................G....TAT...A...G.....................................................
Gypr-DAB1-03(c)                  ...............G....TAT...A...G.....................................................
Gypr-DAB1-04(c)                  ...............G....TAT...A...G.....................................T...............
Gypr-DAB1-01       ...A.........-.............................A.....................................G....A.A..C.A.....
Gypr-DAB1-01c                    .............................A.....................................G....G.A..C.A.....
Gypr-DAB1-01b                    ...............T...........A.....................................G....A.A..C.A.....

Babo-DAB1          ...A.....T...-...  ..A...........TT....ATG...AG..G.........................G......G......A..A.......
Bain-DAB1-06       ...A.T...T...T...  TGA....G......TT....TAT...AA.............G...........CT...G......G....A....A.G...AT
Bain-DAB1          ...A.T...T...T...  TGA....G......TT....GT....A....C...T.....G...........CT...G......G....A.....C.....G
Cyca-DAB1          ........T.C..T...  ..A....T......C.....T.G...A.............................T...G......G....A.A..........
Cyca-DAB2          ............T...  ..A...........CG......G...AA.............................T...G......G................
```

```
Gypr-DAB1-16(g)     CGATCCCAACCGTCTGCAGCAAGAGAGAGCTCAGGTGGAGACAAT
Gypr-DAB1-14(g)     A.........T.........AT..................GTA
Gypr-DAB1-15(g)     A.........T.........AT..................GTA
Gypr-DAB1-13(g)     A.........T.........AT..................GTA
Gypr-DAB1-12(c)     A.........T.........AT..........C.......G.
Gypr-DAB1-11(g)     ....A......T...................G.T.......G.T.
Gypr-DAB1-09(c)     ......................AT..................GTA
Gypr-DAB1-10(c)     ......................AT..................GTA
Gypr-DAB1-08(g)     .A..A......T...................G.T.......G.TA
Gypr-DAB1-07(g)     ............................G.........G.T.
Gypr-DAB1-05(g)     ....A...T.G....................G.........G.G.
Gypr-DAB1-06(g)     ....A......T.........AC.......G...........T.
Gypr-DAB1-02d(g)    A.........T.............................GTA
Gypr-DAB1-02        A.........T.............................GTA
Gypr-DAB1-03(c)     A.........T.............................GTA
Gypr-DAB1-04(c)     A.........T.............................GTA
Gypr-DAB1-01        ......................T...........C..G....TA
Gypr-DAB1-01c       ......................T...........C..G....TA
Gypr-DAB1-01b       ......................T...........C..G....TA

Babo-DAB1           ....A..TT.TT........G.......A...T........G.TA
Bain-DAB1-06        A.........T...........TG..................T.
Bain-DAB1           A......................AC..................TA
Cyca-DAB1           A.........AT............................G.TA
Cyca-DAB2           A...A...C.G.............................CT..T.
```

Figure A3.1: *Gymnocypris* and comparative species *DAB1* locus group sequence alignment. Variable sites in clones for which there is only one sequence will need to be confirmed (refer to Table 3.3). "G" represents sample of genomic origin. "C" represents sample of cDNA origin.

Intron 1

```
Cyca-DAB3          GCAAGTAAAATGAAACTGAGCGAATTTAATGCCTTGATTCAAGTTTTGATTCAAGTTT--TAAATATTTTTATTATTTGATATTCTGAATATA--CTGA
Cyca-DAB4          ..........AAG---..........----------------------------------.....................AA......TA....
Bain-DAB3          ............G...A............-....G-..AA..AA..AATAAAT.AAAACT.GTCA..A.AAG..T..AAT-TCA---------------
Babo-DAB3          .......C............AC.............T...A..AA.G.ATAAGTTT.AA--.G..--...........GATAT.GG..TTGC.GATA.TG


Cyca-DAB3          ATAAACTGGATTTTTATTATTGATATTCATGTTGTTGATATTGTC-ATTGCTGTTAATTCAGTTTATTTGCTTTTACAAAATATTTACATTAATATTCC
Cyca-DAB4          ....-........G.............TG.................-....G..C................T......T.............G.......
Bain-DAB3          ----------.............---------...-.....CT-.............GG...G...................G..G..G........
Babo-DAB3          TA-----------------------------------....CTG.............GG...G...................G..G..G........
Gypr-DAB3-02(g)    ------------------------------------------------------------.....................G..G..G........


Cyca-DAB3          AGAAAGCAATATGAGCC----TTGATAACTAATAGTCTGTCTCTCTGGCTGCTTTCACACTCGGTTCGATTGCTCGA--------CTACTCGCCCTCCT
Cyca-DAB4          .................A---..A...TT.G.C-.C.......--..........T.....T...............TTGCTCGA.............C
Bain-DAB3          .......G......TTAAGATAA.T....AGT.GTGG...T....................T....T..........TCGATCGA..G.G..T...T.C
Babo-DAB3          .......G......TTAAGATAA.T...TAGT.GTGG...T.............C......T..A............CGGATCGA..G.G..T...T.C
Gypr-DAB3-02(g)    .......G......T.AAGATAA.C....ATT.GAAGC..T....................T.........T.....GTGCTCGAT.G.C..T...T.C


Cyca-DAB3          CTGCTCCGCGGTGCATCGTCATCAGCCATCTATGCGG--CACGTGT--ATAGCCTATAGATATATATACGTACATATCTACGGTTTAGC-GCTGTTGTT
Cyca-DAB4          .............GC...G.-----T....AGC.T..--T.GC...TC.....TC.................T.......T.......G-.........
Bain-DAB3          .......C-.....G.-.A.----.T....AGA.T.TGTTT.C.AACGCA.A.AC.G.TCGGC.G..T..CG..CGCG.CT.T.CGGT.-.TG.A...-
Babo-DAB3          ........-C.-..GA-...-----.T.G...CACGAT------CA.CGG------CG.TCATC-GGCG...G.G.C.G.TT.T------C..G.A.A.-
Gypr-DAB3-02()g    ........T.....G-----TGGC.T....AGA.T.TTT.C.-------GCCATC.G.TCATC-GGCG...G.G.C.G.TT.T..GTCG--TG.A...A


Cyca-DAB3          AGTGAACAACTATGCGGGAGACCGTTGCTCTCTTCAATTTTGTGCTTTGGATCTTCGTACTAAATCACGCTATTGGTGATTCGTTTAAACTGTGATTTG
Cyca-DAB4          ........GA.............-........C.GC........------.C...G.C.T-------------------------------------
Bain-DAB3          .A.--.A.GG.TCAAAA..AG.GC...A.T..-.GCTGAAGAGTA.CA...ATGAG-------TA...AT.C.CT.....-GTAGCTTGA-AGTCACGC
Babo-DAB3          .A..C.A..G.TCAAAA..ACGG..CCG.T..-.GCTGAAAA..-.CG...CTGGA--------....-T.C.CT......GCAGCTCGTCAGTCACGC
Gypr-DAB3-02(g)    .-..C.A.GG.TCAAAA..ACG...CAA.T...-G.TGAAGA..A.CA...ATGAGA..AAGTTA...AT.T..A......GCAGCGCGTCAGTCACGC


Cyca-DAB3          CCAGGAAACTGAATTAGTAAAACTAAATCAGT---TTTACAGAATCATCAATAGCGCTTCAGTTCCGTGATTTTTAGTCTGACCGCGTTCACATCAGAG
Cyca-DAB4          ..TT..T...A...C.-CGG...CG..C.GC.---G..G-------------------------.......--....C..G..............
Bain-DAB3          TTGTC..GAAA..C-.CA....TACGT.TTT----GG.CGT-CG........C...G...G.............--.A..C......C......-.....
Babo-DAB3          TTGTC.CGAGAG.CACA..CCGGAC.-.TTTACCGGA.CATACG......GCC...GC..G.............--.-----------......-.....
Gypr-DAB3-02(g)    TTGTC..GAGA.GCC.CA....TAC.G.TTTACCGGA.CGTACG........C...G...G.............A.--..C..........C........
```

```
Cyca-DAB3         TAGAACCGAACCGGAGTCCGTCTGAACCGTAATCAGACCAGACCACGCGTTTCAGGCGGACTC---TTCTCTCTGAAGCTCTAACACTGTGT--TTATG
Cyca-DAB4         C..................A.....................T..............---..TA.....................--.....
Bain-DAB3         .C.....AG...A..----------..AG-..............C.C.G.CT......GGG...AG....GCT..---..T......TC..G..
Babo-DAB3         ........G.A.C.G----------...AG----------ACT......C...G-C.......GGG...AG.G...........T......TT..G..
Gypr-DAB3-02(g)   ........T.........TCA.......A..C.........AA.....C............GGG...AG......T............TT.....

                        Exon 2
Cyca-DAB3         TTTCAGCTCATGGATATTACGAGTACATGACGTTTGACTGCGTCTACAGCGCCAGTGATTACAGTGATATGGTGTATCTTCTGTCATATTCCTTCAATC
Cyca-DAB4         ........G.................C..T..A...A...........A.............................G.....CT..........
Bain-DAB3         .......GG.............G.ATG..T.CC...A.................A..C.....A..............G.........T.......G
Babo-DAB3         .......G............TC....GG..T.CC...A............A.............................G.................A
Gypr-DAB3-02(g)   ........G............TC....CA..T..CA..A............A.............G..............................
Gypr-DAB3-01(c)   ------..G...................TGA......A............A.........................T....AC...............A
Gypr-DAB3-02(c)   ------..G............TC....CA..T..CA..A............A.............G..............................
Gypr-DAB3-03(c)   ------..G............AC.....C..T.AG...A............A.......C.................T....GA...............G
Gypr-DAB3-03e(c)  ------..G............AC.....C..T.AG...A............A.......C.................T....GA...............G
Gypr-DAB3-04(c)   ------.GG.................C.T...A................A.......C.................T....GC..T..T..........A

Cyca-DAB3         AAGGCCTGGATGCACAGTTCAACAGCTCTGTGGGGAAGTTTGTGGGGTACACTGAGAAAGGAATGAAAGATGCAGAGTACTGGAACAATAATCCGGCAG
Cyca-DAB4         ...T.G......T...A.G...........C.AT.....G.................G.....G.....T.........A...TC.....AG.....AGC.
Bain-DAB3         ...CTA..C.........A.......A.A.........................G.GT....G.........................C....A..ACT
Babo-DAB3         ...TTG.....AT.............A.C.............................C.....G.....TT.......A...TC.....A.........CA
Gypr-DAB3-02(g)   ...CTG.T...CT.....A.......A.A.....AG.......T....C.....C.C.....G.....T.........A...TC.....A..C.....CA
Gypr-DAB3-01(c)   ...TTG......TG....A.......A.A.............................G.....T.....T.........A...TC.....AG...A...CT
Gypr-DAB3-02(c)   ...CTG.T...CT.....A.......A.A.....AG............C.....C.C.....G.....T.........A...TC.....A..C.....CA
Gypr-DAB3-03(c)   ...T.G.....CT..T.....................................C.G.....G.....TT.......A...TC.....AG...A...CT
Gypr-DAB3-03e(c)  ...T.G.....CT..T.....................................C.G.....G.....TT.......A...TC.G...AG...A...CT
Gypr-DAB3-04(c)   ...T.G......T.........................................G.....G.....T.........A..........AG...A...CA

Cyca-DAB3         AACTGCAGCAGAGGAGAGCTGAGGTGGACACATA
Cyca-DAB4         TT......G.TCT..A.A..TCA...........
Bain-DAB3         .CA.....GGACT..A..G...C.....G.G..T
Babo-DAB3         TCA........CT..A..G...A.......G..T
Gypr-DAB3-02(g)   TCA........CT..A....A.C.......G..T
Gypr-DAB3-01(c)   TCA........CA..A.....CA..........T
Gypr-DAB3-02(c)   TCA........CT..A....A.C.......G..T
Gypr-DAB3-03(c)   TCA.A......GA..A.....CA..........T
Gypr-DAB3-03e(c)  TCA.A......GA..A.....CA..........T
Gypr-DAB3-04(c)   TCA.A......CT..A....A.T.......G..T
```

Figure A3.2: *Gymnocypris* and comparative species *DAB 3* locus group sequence alignment. Variable sites in clones for which there is only one sequence will need to be confirmed (refer to Table 3.3). "G" represents samples of genomic origin. "C" represents samples of cDNA origin.