Applications of Microsatellite Markers to Genetic Management of Carps in Aquaculture

By

Almas Ara Gheyas

A thesis submitted for the Degree of Doctor of Philosophy

Institute of Aquaculture
University of Stirling
Stirling, Scotland, UK

February, 2006

This thesis is dedicated to my beloved husband
Md. Mofakkarul Islam

Declaration

I declare that this thesis has been compiled entirely by me based on my own investigation. It has not been submitted for any other degrees. All information from other sources or any assistance received has been duly acknowledged.

Almas Ara Gheyas
February, 2006

Signature of Candidate _____

Signature of First Supervisor _____

Signature of Second Supervisor _____

ABSTRACT

Carp aquaculture in South Asia suffers severely from a lack of genetic management, which has eroded the genetic quality of both captive and wild populations. Use of molecular markers, especially microsatellites, has revolutionized genetic management of hatchery stocks through its ability to detect kinship between individuals and hence in controlling level of inbreeding and loss of genetic diversity. In the present PhD work, microsatellite markers were applied to breeding programmes for silver carp (*Hypophthalmichthys molitrix*) and common carp (*Cyprinus carpio*) to study different genetic management aspects and new markers were generated from rohu (*Labeo rohita*).

A set of newly isolated microsatellite markers from silver carp were characterized and two pentaplex PCR reactions were optimized to enable rapid genotyping of large number of individuals at 10 microsatellite loci. The utility of these markers in parentage, sibship and relatedness analysis were assessed by applying them to groups of fish with known relationship. These markers were used for parentage analysis in a breeding programme designed to estimate heritability of harvest weight and length in silver carp. Full- and half-sib families were created in three sets of partly factorial mating and all the families from each set were reared in communal ponds from very early life stages. With ten microsatellites 96.3% of the offspring could be assigned to a single family. Heritability estimates were found to be 0.65 ± 0.13 for weight and 0.50 ± 0.13 for length. High estimates of h^2 suggested that this population should respond rapidly to selection for increased harvest size.

Microsatellite markers were also applied to monitor the early stages of a mass selection programme in common carp (*Cyprinus carpio*). The selection was initiated from a base

population synthesized from six different stocks. The selected individuals were divided to create two separate lines. The aims of this study were to monitor whether the stocks were represented in the intended proportions in the F_1 selected populations, to investigate the relative contribution of families and its impact on effective population size and to identify any loss of molecular genetic variation. Five highly polymorphic microsatellites were used for parentage analysis of the selected fish to track stock and family contribution. Overall, large perturbations were observed in the relative contributions of two major stocks. Family contribution was also highly variable, causing the N_e to drop to below half the census size. A loss of 6.9%-12.2% of microsatellite alleles was observed but loss of heterozygosity was not very prominent. The replicate lines showed significant differences in allelic distribution after the first generation of selection, but not in genotypic distribution.

Finally, 52 microsatellite markers were isolated from a partial genomic library of rohu using a selective hybridization protocol. Characterization of these markers resulted in 36 polymorphic loci, which will be useful in future work on conservation and management of both wild and captive rohu populations.

Acknowledgements

I take this as an opportunity to acknowledge the help and support received from a number of people/organisation during the course of this PhD study. First, I would like to extend my grateful thanks and sincere gratitude to my supervisors Dr. David J. Penman and Professor Brendan J. McAndrew for their scholastic supervision, constant support and ready availability. It was indeed a great privilege to work with them, to learn from them and to receive their expert suggestions whenever I needed.

I am very grateful to the Commonwealth Commission for financing my studies, without which it would have not been possible for me to come to UK for pursuing PhD studies. I also offer my sincere thanks to BFWG Charitable Foundation for providing me with supplementary grant after the end of my Commonwealth Scholarship. This helped me sustain myself during the very hectic period of writing this thesis.

My special thanks are due for Dr. John B. Taggart for all his patience and time in helping me with the lab techniques. I also extend my grateful thanks to Dr. John Woolliams of Roslin Institute for the help with major statistical analyses.

The breeding and rearing part of one of my major experiments was conducted in Bangladesh. I wish to thank Mr. Abdus Sattar, Mr Topan K. Das and other staff of the North-West Fisheries Resource Development and Management Project (NFRDMP), Parbatipur, Bangladesh for taking care of that part in collaboration with Dr. David Penman. I also gratefully acknowledge the assistance provided by Dr. Y. Basavaraju of the Karnataka Veterinary, Animal and Fisheries Science University, Bangalore, India and his staff in diligently sending the tissue samples and the necessary information for two other major experiments. I am grateful to Professor Shamsul Alam of Bangladesh Agricultural University for his generosity in providing me with the rohu samples.

Many of the staffs of the Institute of Aquaculture have been ready source of support and help, most notably Mrs. Ann Gilmour, Mrs. Cathryn Dickson, Mr. Stephen Powell, Mrs. Elizabeth Stenhouse and Dr. Mike Leaver. All their help and support were greatly appreciated. I would also like to extend my thanks to Mrs. Kate Howie for giving me advice on some of the statistical issues.

During my studies in the University of Stirling, I have made a number of good friends to whom I owe many of the enjoyable moments. I would specially like to mention Dr. Tariq Ezaz, Dr. Sayema Sayeed, Marine Herlin, Karen Frake, Anu Frank-Lawale, Amid Derayat and Jose Mota-Velasco for creating a very friendly atmosphere around me and for providing me with mental support all through.

My family members have been the endless source of support during my PhD study. I very gratefully acknowledge the constant encouragement, blessings and love of my parents and parents-in-law, my sisters, brother, sisters-in-law and brothers-in-law. Without their encouragement and blessings completing this PhD would have been very difficult.

The contribution of my beloved husband in my PhD studies is beyond my ability to express. I can only say that it would have been extremely difficult, if not impossible, to carry on this PhD study without his enormous sacrifice yet endless support and love.

TABLE OF CONTENTS

Chapter 3. Characterization of silver carp (*Hypophthalmichthys molitrix*) microsatellites and their utility in kinship analysis

Chapter 4. Estimation of heritability of harvest traits in silver carp

Chapter 5. Monitoring a selective breeding programme in common carp (*Cyprinus carpio*) using microsatellite markers

Chapter 7. General Discussion

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF SPECIES NAMES

Common name	Scientific name
Abalone	*Haliotis sp.*
African catfish	*Clarias gariepinus*
Atlantic salmon	*Salmo salar*
Bighead carp	*Aristichthys nobilis*
Black bream	*Acanthopagrus butcheri*
Black carp	*Mylopharyngodon piceus*
Black tiger shrimp	*Penaeus monodon*
Brown trout	*Salmo trutta*
Catla	*Catla catla*
Channel catfish	*Ictalurus punctatus*
Chinook salmon	*Oncorhynchus tshawytscha*
Chum salmon	*Oncorhynchus keta*
Cod	*Gadus morhua*
Coho salmon	*Oncorhynchus kisutch*
Common carp	*Cyprinus carpio*
Gilthead sea bream	*Sparus aurata*
Grass carp	*Ctenopharyngodon idella*
Haplochromine cichlid	*Prognathochromis perrieri*
Japanese flounder	*Paralichthys olivaceus*
Kuruma shrimp	*Penaeus japonicus*
Mrigal	*Cirrhinus mrigala*
Nile tilapia	*Oreochromis niloticus*
Pink salmon	*Oncorhynchus gorbuscha*
Rainbow trout	*Oncorhynchus mykiss*
Red sea bream	*Pagrus major*
Rohu	*Labeo rohita*
Sea bass	*Dicentrarchus labrax*
Sea urchin	*Strongylocentrus intermedius*
Senegalese sole	*Solea senegalensis*
Silver barb	*Barbonymus gonionotus*
Silver carp	*Hypophthalmichthys molitrix*
Silver crucian carp	*Carassius langsdorfi*
Steelhead	*Oncorhynchus mykiss*
Toothfish	*Dissostichus eleginoides*
Turbot	*Scophthalmus maximus*
Wild mice	*Mus musculus*
Zebrafish	*Danio rerio*

CHAPTER I

GENERAL INTRODUCTION: GENETIC MANAGEMENT IN AQUACULTURE AND MOLECULAR MARKERS

1.1 Aquaculture and genetic management: an overview

Aquaculture has flourished rapidly as an industry and livelihood activity in many parts of the world, particularly in Asia in recent decades, to boost the production of fish for food and game (Liao 1997). It has been identified as the fastest growing food production system, increasing at an average annual rate of 8.9% compared to only 1.2% for capture fisheries and 2.8% for total livestock meat production (FAO 2004). Carps, salmonids, and tilapia are the most widely cultured fishes. Due to overfishing, destruction of fish habitat, pollution and blockage of migratory routes many wild fish stocks have declined drastically, reducing the production from nature (Waples 1999). Aquaculture is being used to compensate for the shortage of production from capture fisheries. Moreover, hatchery produced juvenile fish are being released to open water bodies to rejuvenate declining natural stocks.

The rapid expansion of commercial aquaculture, however, is incurring a new array of problems (Ferguson 1995). Some of these problems are very obvious and quick in their effects, for instance, outbreak of diseases (Johnsen and Jensen 1994; Rogne 1995) or pollution of water bodies through the release of untreated effluents from aquaculture farms (Hakanson *et al.* 1988; GESAMP 1991). There are, however, covert problems which can potentially destroy the sustainability of the whole aquaculture business but their effects are not manifest immediately. Such problems arise due to a lack of a proper genetic management of farm stocks. Lack of genetic management alters the genetic constitution of a culture stock, often leading to poor performance in terms of growth and survival. It is, however, not only the sustainability of aquaculture that is at stake due to the absence of genetic management, but also the integrity of the wild aquatic stock. Natural stocks are put at risk when genetically different farm fish gain entry into the wild either by accidental escape or by humans deliberately releasing them for stock

enhancement, and they then interact with wild fish. The detrimental effects of hatchery fish on wild stock have been reported in many species; some examples are steelhead, *Oncorhynchus mykiss* (Chilcote *et al.* 1986), brown trout, *Salmo trutta* (Hansen and Loeschcke 1996), Atlantic salmon, *S. salar* (Clifford *et al.* 1998), coho salmon, *O. kisutch* and chinook salmon, *O. tshawytscha* (Noakes *et al.* 2000).

As a consequence of numerous reports on the poor performance of hatchery reared fishes and the impacts of release of hatchery fish in to the wild, genetic issues in aquaculture are now receiving due attention. Scientists are now using all the available modern techniques at their disposal to understand the mechanisms that lead to genetic changes in aquaculture and to design proper management techniques, suitable for particular situations. Molecular markers have proven to be extremely important tools in this endeavour. Different types of molecular markers are now available and are finding applications in aquaculture genetics.

The present PhD used microsatellite markers to study different aspects of genetic management in carp aquaculture in South Asia (India and Bangladesh). To provide thorough background information, this chapter reviews relevant literature on the following major aspects:

1. Importance of genetic management in aquaculture
2. Available molecular markers and their potential roles in genetic management
3. Microsatellite markers (special highlights)
4. Applications of microsatellite markers in aquaculture
5. Carp culture in South Asia with special highlights on Bangladesh

1.2 Genetic management in aquaculture

1.2.1 Aquaculture and genetic changes

The importance of genetic management in aquaculture can be fully appreciated if the causes of genetic changes in aquaculture and their potential impacts are properly understood. A number of genetic processes are responsible for bringing about genetic changes in aquaculture stocks. These include: (i) genetic drift and inbreeding (ii) inadvertent selection of broodstock, (iii) relaxation of natural selection and natural selection in captivity and (iv) artificial selection (Busack and Currens 1995; Waples 1999; Mignon-Grasteau *et al.* 2005). Apart from these factors, deliberate genetic manipulation events also create genetically altered individuals or stocks.

Genetic drift and inbreeding are characteristic features of small and closed populations such as those maintained in hatcheries. These are called dispersive processes as they lead to random variations in gene frequencies. Genetic drift is the change in gene frequency arising from an error in sampling of gametes. The gametes that transmit genes to the next generation carry a sample of the genes from the parent generation, and if the sample is not large enough the gene frequencies are liable to change between one generation and the next. As a consequence the genetic diversity is reduced. Even one generation of artificial spawning and hatchery rearing can cause large changes in the genetic make-up of a population, especially by decreasing the overall level of genetic variation (Allendorf and Ryman 1987; Cross 1999).

Inbreeding means mating of individuals that are related to each other by ancestry. Inbreeding, by itself, does not lead to changes in gene frequency or variety of genes in a population (Falconer and Mckay 1996). This however, increases the frequency of homozygotes at the expense of heterozygotes in subsequent generations because related

individuals share common alleles. This leads to changes in frequency of phenotypes in the population. If selection then acts on these phenotypes, allele frequencies can also change (Busack and Currens 1995). Increased homozygosity may result in a reduction in the mean phenotypic values of fitness-related traits such as reproductive capacity or physiological efficiency. This phenomenon is known inbreeding depression, which may result from two major causes. First, if heterozygotes perform better than homozygotes, then the decrease in heterozygosity will lead to a decrease in performance. Second, deleterious alleles of many traits are recessive in nature. Increasing homozygosity, therefore, increases the chances of expression of deleterious recessive alleles (Tave 1993; Falconer and Mackay 1996). The combination of inbreeding and genetic drift can quickly destroy a population's genetic variance (Tave 1993). The extreme effect of these two processes is the fixation of one allele in the population and complete loss of other alleles for that locus (Falconer and Mackay 1996).

Inadvertent or unintentional selection occurring as a part of routine operations often causes large genetic changes in hatchery populations. Unintentional selection may occur from many different sources, such as from replacement of the broodstock in a hatchery with only a few fingerlings after the best ones have been sold, not being able to select breeders randomly due to netting bias, choosing few but the most fecund females or the biggest males for seed production, etc. (Campton 1995; Penman and McAndrew 1998).

Relaxation of natural selection means a reduction of selection pressure that would normally act under wild conditions. It applies to traits that are important in nature but not in captivity, such as food finding, seasonal reproduction, body colour and predator avoidance (Mignon-Grasteau *et al.* 2005). Domestic animals can thus be more variable for these traits than their wild counterparts (Price and King, 1968). While there is

reduction in the natural selection for certain traits, there is an increased selection for other traits which are necessary for cultured animals to excel in their new captive condition. This has been defined by the concept of 'local adaptation' (Carvalho 1993; Busack and Currens 1995). Local adaptation is a process that increases the frequency of those traits which enhance the survival or reproductive success of individuals in a particular environment. Since a hatchery differs in many ways from the natural environment (Waples 1999), the gene pool of cultured fish is likely to become very distinct from that of wild fish (Skaala *et al.* 1990).

Artificial selection involves deliberate selection of broodfish for desirable traits such as higher growth rate, better disease resistance etc. Ferguson (1995) explains that such a selection programme changes the genetic make-up of the base population by replacing "negative" alleles with "positive" ones leading to reduction in genetic variability. This reduction in genetic variation is not deleterious in commercial aquaculture as long as the strains have the desired combination of genes and phenotypes. However, a potentially harmful consequence of directional artificial selection is inbreeding. The use of only a small fraction of possible spawners where selection is intense increases the probability of mating between relatives leading to significant reduction in stock productivity.

Unlike inbreeding and genetic drift, artificial selection and natural selection modify gene frequencies in a given and foreseeable direction (Mignon-Grasteau *et al.* 2005).

1.2.2 What is genetic management and how to apply this in aquaculture?

The spectrum and goals of genetic management varies with the purpose of the aquaculture or hatchery operation, that is, whether fish are being produced commercially as food fish or for stocking in open water bodies for rehabilitation of a natural population. The major components of genetic management in different types of

aquaculture are shown diagrammatically in Figure 1.1. The diagram shows that irrespective of the purpose of the aquaculture programme, two components of genetic management are common: to maintain genetic diversity and to monitor the genetic changes in the population so that mitigating measures can be taken (Hedrick *et al.* 1986; Allendorf and Ryman 1987; O'Connell and Wright 1997). Maintaining genetic diversity is the most critical component because this provides a population the power to evolve in response to environmental changes. Also the fitness of a population is believed to be directly dependent on genetic diversity (Beacham 1991; Carvalho 1993; Reed and Frankam 2003). The genetic changes that take place in an aquaculture/hatchery environment lead to the loss of genetic diversity. The prime goal of any genetic management programme, therefore, is to reduce the effect of factors causing genetic changes as much as possible. Some genetic changes in a captive population, however, cannot be eliminated entirely no matter what measures are taken (Busack and Currens 1995; Waples 1999).

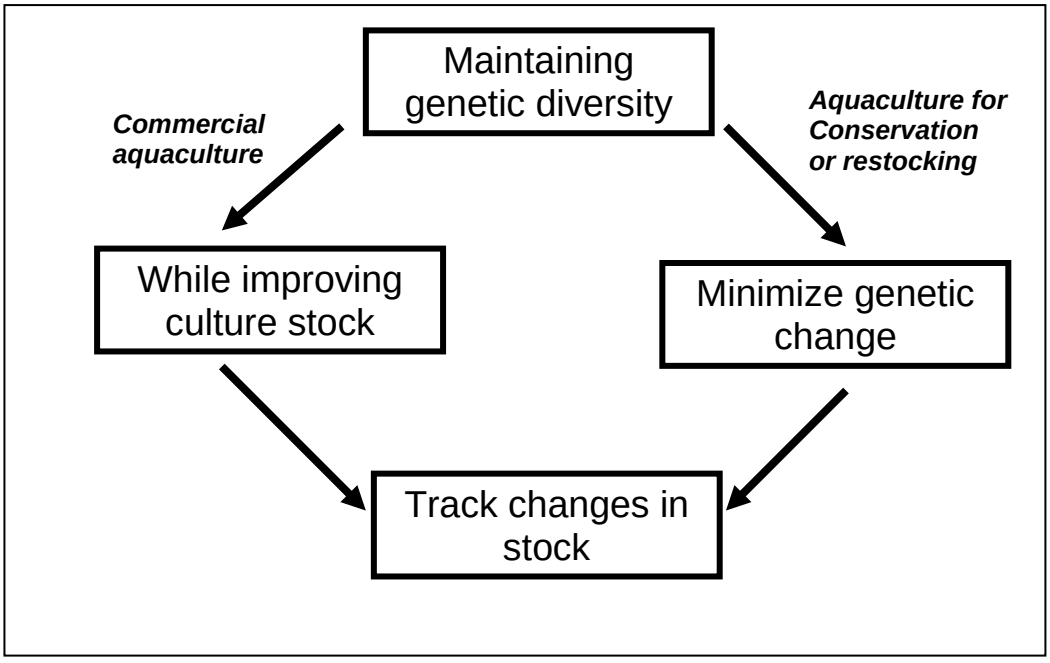

Figure 1.1: Diagram showing the components of genetic management in different types of aquaculture practice

A commercial aquaculture breeding programme should aim at increasing the production mainly by selective breeding. Genetic management in such culture practices should be applied such that genetic improvement can be performed without drastic reduction of genetic variation through inbreeding and random drift. Ferguson (1995) suggests that in this case it is important to start with a founder population with a broad genetic base so that it allows choosing the most desirable phenotypic traits. Besides this, it is extremely important to maintain a large effective population size (N_e). A reduction in N_e can seriously affect a population's biological potential, can adversely affect productivity, and can ruin the opportunity to further improve the stock via selection (Tave 1993). To maintain a large N_e, as large a number of fish as possible should be spawned, the sex ratio of the spawners should be kept close to 1:1 and the variance of family size should be reduced. Maintaining detailed pedigree information should constitute another important aspect of genetic management in selective breeding programmes to avoid crossing between closely related fishes.

When selective breeding is not the goal in a hatchery, care must be taken to prevent any unintentional selection through a biased choice of brood fish. Tave (1993) describes this as a "no selection" approach. He suggested the following ways to prevent unintentional selection: (i) breeding fish over the entire spawning season, (ii) spawning broodfish of all sizes, (iii) breeding as many fish as possible, and (iv) not culling the slow growers or those with less desirable qualities, i.e. choosing a representative sample from broodfish population. Other management approaches such as maintaining a large N_e and avoiding inbreeding are the same as discussed above.

If the fish are being produced for restocking into the wild, the major focus of management should be to keep the genetic composition and the level of variability of

the hatchery stock representative of the local natural population (Allendorf and Ryman 1987; Cross 2000) so that the risk of cultured fish on wild population can be minimized. Many authors have described the possible genetic impacts of mixing cultured fish with wild ones when they differ in their genetic composition (Hindar *et al.* 1991; Ryman *et al.* 1995; Rhymer and Simberloff 1996; Cross 2000; Johnson 2000). The impacts are mainly of two broad kinds: (1) introgression of exogenous genetic material and (2) reduction of effective population size.

The major threat associated with the **introgression of exogenous genetic material** is the outbreeding depression which means that the fitness of the offspring declines as the genetic differences between parents increases (Templeton 1986; Emlen 1991). It is assumed that the local adaptation of a population is achieved through a particular arrangement of alleles at different loci, termed as co-adapted gene complexes (Cross 2000). Interbreeding between two genetically distinct populations may lead to a breakdown of these complexes, resulting in a reduced fitness. In this way the native genetic pool is eroded with the ultimate effect of weakening or even loss of the natural population.

The **effective population size (N_e)** may be reduced due to a number of reasons e.g. unequal sex ratio, non-random mating, overlapping generations and unequal number of progeny per family (Tave 1993). A stock enhancement programme increases the variance in family size since the fish bred in captivity contribute many more offspring in the next generation than the fish that reproduced in the wild. The result is a drastic reduction in the effective population size, even though the census size is increased (Ryman *et al.* 1995).

Considering the above impacts, rehabilitational aquaculture should incorporate the following features in the management practices (after Skibinski 1998 and Cross 2000):

1. Monitoring of released fish using molecular markers.
2. Choosing that strain of hatchery fish which is representative of the local population.
3. Minimizing genetic change during captive breeding. No strain manipulation should be undertaken and other agents of change should be minimized.
4. Broodstock should be collected randomly in numbers sufficient to avoid inbreeding effects and all of them should participate in spawning, preferably using single pair mating. Population numbers in the 100s or 1000s rather than 10s appear appropriate for this purpose.

1.3. Molecular markers and their potential roles in genetic management

Molecular markers have been used in many different aspects of genetic management in aquaculture. Their roles in aquaculture have been reviewed by Ferguson (1995), Magoulas (1998), Dunham (2004), and Liu and Cordes (2004). Various roles of molecular markers include:

- Strain or species identification
- Detection of inter and intra-specific hybridization
- Parentage and kinship analysis
- Assessment of parental contribution in mass spawning
- Estimation of N_e and level of inbreeding
- Preventing inbreeding
- Mapping of quantitative trait loci (QTLs) and selective breeding

In the management of rehabilitational aquaculture, genetic markers can play the following important roles (after Cross 2000):

- Comparison of farmed strains and wild populations
- Choice of donor population
- Detection of genetic changes in hatchery reared fishes over generations
- Monitoring the impact of reared animals after release to the wild

1.3.1 Different types of molecular markers

The need to detect genetic variation has made the search for novel marker systems a continued process. The major available molecular markers are: isozymes, mitochondrial DNA (mtDNA), restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites, expressed sequence tags (ESTs) and single nucleotide polymorphism (SNP). The markers have been classified into two categories: **type I** are markers associated with genes of known function, while **type II** markers are associated with anonymous genomic segments (O'Brien 1991). Different markers have found their usage in different aspects of aquaculture genetics. Type and characteristics of the marker, ease and expense of application, abundance in the genome and polymorphic information content (PIC) are some of the major factors that make one molecular marker more useful than others under different situations. PIC refers to the value of a marker for detecting polymorphism in a population. It depends on the number of detectable alleles and the distribution of their frequencies (Liu and Cordes 2004).

Except microsatellites, all the different types of molecular markers thus far used in aquaculture are discussed in the following sub-sections. Since the present research work

used microsatellites, two full sections (1.4 and 1.5) are dedicated to describe the characteristics, principles and applications of these markers.

1.3.1.1 Isozymes

Isozymes are multiple molecular forms of individual enzymes. These multiple forms can be products of different alleles at a single locus –allozymes- or products of different loci where multiple copies of genes make the same enzyme or enzyme subunits. Isozymes are separated in an electric field passed through a matrix, such as starch, cellulose acetate or polyacrylamide, based on their size, shape and charge and are visualized with specific histochemical stains (Dunham 2004). Technically easy application, requirement of simple and inexpensive facilities and rapid processing of large numbers of samples have made isozymes highly popular (Park and Moran 1995). Another major advantage is that isozymes are considered Type I markers and the genetic variation measured through them may actually affect performance. Isozyme variation has been found to show association with growth, disease resistance, temperature tolerance, developmental speed, and salinity tolerance in fish (Dunham 2004). In aquaculture, isozymes have been applied for tracking inbreeding, stock identification, parentage analysis and linkage mapping (Liu and Cordes 2004). A number of disadvantages, however, have limited their use: (i) sample collection for their assay often requires sacrifice of the animal which is not desirable when dealing with, for instance, endangered species; (ii) a large quantity of tissue is required, making them inapplicable for small animals (e.g. larvae); (iii) isozymes reflect such a small portion of the genome that sufficient variation may not exist in assayable loci to discriminate between recently diverged or inbred hatchery populations; (iv) even if variation exists in the underlying DNA sequence, that may not be represented at the protein level, reducing the level of detectable variation. This happens because some changes in nucleotide

sequence do not change the encoded polypeptide (silent substitution) and some polypeptide changes do not alter the mobility of the protein in an electrophoretic gel. Heterozygote deficiency also occasionally appears from null alleles (total inactivation of the enzyme due to changes in DNA sequence) (Park and Moran 1995; Magoulas 1998; Liu and Cordes 2004).

1.3.1.2 Mitochondrial DNA (mtDNA)

Sequence divergence has been observed to accumulate more rapidly in mitochondrial than in nuclear DNA (Brown, 1985). This has been attributed to a faster mutation rate in mtDNA resulting from a lack of repair mechanism during replication and smaller effective population size due to strictly maternal inheritance of the haploid mitochondrial genome (Liu and Cordes 2004). High genetic variance compared to isozymes made mtDNA a very popular marker and its use dominated the studies regarding phylogeny and population differentiation during the 1980s (Magoulas 1998). However, maternal inheritance of mtDNA raises concerns that it may fail to reflect the true phylogenies and population structures if gender-biased migration or introgression have occurred (Birky *et al.* 1989; Chow and Kishino 1995).

1.3.1.3 Restriction fragment length polymorphism (RFLP)

Digestion of DNA with a set of restriction enzymes creates fragments of different sizes. Such length polymorphism (called RFLP) results from a base substitution that causes the gain or loss of a restriction site, or from insertions, deletions or DNA rearrangements at or between the restriction sites. This technique can detect variations among individuals, populations and species and hence has found its use as a molecular marker. Traditionally, Southern blot analysis was used to detect the length polymorphism, where the digested fragments upon electrophoresis through an agarose

Table1.1: Types of molecular markers, their characteristics, and potential applications (from Liu and Cordes 2004).

Marker	Mode of inheritance	Type	Likely allele numbers	Polymorphism or power	Major application
Isozyme	Mendelian, codominant	Type I	2-6	Low	Linkage mapping, population studies
mtDNA	Maternal inheritance	--	Multiple haplotypes		Maternal lineage
RFLP	Mendelian, codominant	Type I or II	2	Low	Linkage mapping
RAPD	Mendelian, dominant	Type II	2	Intermediate	Population studies, hybrid identification
AFLP	Mendelian, dominant	Type II	2	High	Linkage mapping, population studies
Microsatellite	Mendelian, codominant	Mostly Type II	Multiple	High	Linkage mapping, population studies, Paternity analysis
EST	Mendelian, codominant	Type I	2	Low	Linkage, physical and comparative mappings
SNP	Mendelian, codominant	Type I or II	2, but up to 4	High	Linkage mapping, population studies

gel were transferred to a membrane, and visualized by hybridization to specific radioactive or fluorescent probes. Recently, PCR based technique has been used to replace the tedious Southern blot method. The advantages of RFLP markers include their codominant inheritance and easy interpretation and scoring. Their major disadvantage is the low level of polymorphism compared to more recently developed markers. Besides, sequence information required for primers or probes may limit their use (Dunham 2004; Liu and Cordes 2004).

1.3.1.4 Randomly amplified polymorphic DNA (RAPD)

The RAPD technique uses short primers (8-10 bp) of arbitrary sequence to amplify anonymous regions of genomic DNA. The use of short primers and low annealing temperatures (often 36-40^{o}C) allows amplification of multiple products, each product (presumably) representing a different locus. Genetic variations are assessed by the presence or absence of each product. The potential power of RAPD for detection of polymorphism is relatively high, because 5-20 bands are typically produced by a single primer and multiple sets of primers can be used to scan the entire genome. However, each RAPD locus is considered bi-allelic (presence or absence of an amplified product) and hence their PIC values fall below those for microsatellites and SNPs (Liu and Cordes 2004).

Being a PCR-based technique, RAPD is an easy, fast and inexpensive method. No sequence information of the target DNA is required to design the primers. RAPDs have been used in a variety of aspects of aquaculture genetics, such as: species identification, detection of interspecific hybridisation, analysis of population structure, estimation of heterosis in strain crosses and analysis of genetic diversity (Elo *et al.* 1997; Shikano and Taniguchi 2002; Liu and Cordes 2004).

The major shortcomings of this marker are poor reproducibility and ambiguity in the interpretation of results. RAPDs are dominant markers and hence fail to distinguish between homozygous and heterozygous individuals. In addition, amplified products from different regions may have the same length and can lead to a wrong conclusion that the products represent a single locus (Magoulas 1998; Liu and Cordes 2004).

1.3.1.5 Amplified fragment length polymorphism (AFLP)

AFLP is a PCR-based, multi-locus fingerprinting technique that combines the strengths and overcomes the weaknesses of the RFLP and RAPD methods. AFLP method involves four steps: digestion of total genomic DNA, ligation of adapters, PCR amplification of selected fragments and gel analysis. Since the ligated adapters have known sequence, they are used as primer sites. Selective PCR amplification is performed by adding known bases to the $3^{/}$ end of the PCR primers. Since these bases extend past the ligated site and into the DNA fragment, the primer anneals only if the fragment has the correct sequence.

The major strengths of the AFLPs include: high polymorphism, good reproducibility due to high PCR annealing temperatures, and relative economy on a per marker basis. It does not require any prior molecular information and thus is applicable to any species. Its major weakness is the need for specialized equipment or lab facilities. Like RAPDs, AFLP markers are inherited in a dominant fashion, although software packages are now available for co-dominant scoring of AFLP bands (AFLP-Quantar^{TM} Pro Image analysis software by Greelings *et al.* 1999).

AFLP markers have been used for generating high resolution linkage maps, in population genetics studies and analysis of gynogenetics and androgenetics (Dunham 2004; Liu and Cordes 2004).

1.3.1.6 Single nucleotide polymorphism (SNP)

SNP describes polymorphism caused by point mutations that give rise to different alleles containing alternative bases at a given nucleotide position. SNPs can be determined by DNA sequencing, primer extension typing, the designing of allele specific oligos and gene-chip technology. SNPs have received enormous attention in human genetics and several model organisms since they are the most abundant polymorphism in the genome, adaptable to automation, and reveal hidden polymorphism not detected with any other methods. Theoretically, a SNP locus can have up to four alleles, each containing one of the four bases. Practically, however, most SNPs are usually restricted to two alleles (most often either the two pyrimidines C/T or the two purines A/G) and have been regarded as bi-allelic. As a result, their PIC is not as high as multi-allele microsatellite, but this shortcoming is balanced by their great abundance. SNP markers are inherited as co-dominant markers. Despite technological advances, SNP genotyping is still a challenging endeavour and requires specialized equipment, so their use in aquaculture related research is not yet widespread (Liu and Cordes 2004).

1.3.1.7 Expressed sequence tags (ESTs)

ESTs are small pieces of DNA sequence (usually 200-500 bp long) that are generated by sequencing cDNA clones. The principle of EST technique is to sequence part of genes expressed in certain cells, tissues, or organs from different organisms and use them as tags to locate a gene on chromosomal DNA. ESTs are considered biallelic codominant markers. Their polymorphism results from the presence or absence of a transcript. ESTs are powerful tools in identifying genes and analysis of their expressions at different tissues or under different situations. ESTs have been found to be very efficient markers for linkage and physical mapping and for comparative genome

mapping (Reviewed by Liu and Cordes 2004, Dunham 2004). Nonpolymorphic ESTs are used for genetic mapping through creation of radiation hybrid panels. Microsatellites can be found within ESTs which makes them even more useful for gene-mapping research.

1.4 Microsatellite markers

Microsatellites, also known as "Simple Sequence Repeats" (SSR) are tandem arrays of short nucleotide repeats (1-6 base pairs) and are interspersed throughout the genome (Tautz and Renz 1984; Tautz 1989). Except trinucleotide microsatellites, all other types are generally more prevalent in the non-coding regions of the DNA compared to the protein-coding regions (Li *et al.* 2002). Microsatellites have been detected within the genome of every organism so far analysed and are often found at frequencies much higher than would be predicted purely on the grounds of base composition (Hancock 1999). The majority of microsatellites (48-67%) found in many species are dinucleotide (Wang *et al.* 1994) but in primates, mononucleotides, mainly poly A/T tracts, are the most abundant classes of microsatellites (Tóth *et al.* 2000). Mononucleotide microsatellites, however, are unsuitable to be used as markers because they are highly unstable during PCR, making allele scoring difficult or impossible (Hancock 1999).

Although microsatellites are usually considered as evolutionary neutral DNA markers, there is now much evidence that microsatellite sequences also serve functional roles as coding or regulatory elements. Their function has been proven in chromatin organisation, regulation of DNA metabolic processes such as DNA replication, recombination, cell cycle etc. and in regulation of gene activity such as transcription, binding protein and translation (reviewed by Li *et al.* 2002).

Microsatellites may be classified into three families — pure, compound and interrupted repeats (Jarne and Lagoda 1996; Hancock 1999). The pure forms consist of single sequence motif repeats and the compound forms consist of contiguous or adjacent tandem arrays of different motifs. The third type originates when a microsatellite acquires numerous point mutations and as a result the repeat arrays are interspersed with non-repeat sequences. This pattern is also known as cryptic simplicity and is common in many genomes, particularly in higher eukaryotes (Tautz *et al.* 1986; Sarkar *et al.* 1991; Hancock 1995).

1.4.1 Attributes that make microsatellites good genetic markers

Several attributes of microsatellites render them extremely valuable as genetic markers for numerous applications. Firstly, microsatellites are very abundant and dispersed throughout the genome. This makes them especially suitable for genome mapping studies (O'Connell and Wright 1997). On an average, they occur approximately once every 10 kbp (Wright 1993).

Second, microsatellites show extremely high levels of polymorphism that make their PIC values the highest among all the markers (Litt and Luty 1989; Weber and May 1989; Liu and Cordes 2004). This attribute makes them especially attractive in studying genetic variation in species or populations that show low overall variance with conventional markers such as isozymes and mtDNA, for studying inbred and bottlenecked populations often found in hatcheries and in pedigree analysis (O'Connell and Wright 1997).

Third, microsatellite alleles show co-dominance and are inherited in a Mendelian fashion (DeWoody and Avise 2000). This makes them particularly suitable for pedigree analysis and for population studies (O'Connell and Wright 1997).

Fourth, being small in size microsatellites can be amplified by PCR. While isozyme and mtDNA studies require fresh or frozen tissue, only a small amount of blood or fin clip or scale is required for microsatellite analysis. These tissues can be obtained without sacrificing the animals and stored in alcohol (Carvalho and Hauser 1995; Wright and Bentzen 1995).

Finally, microsatellites can be assayed rapidly compared to many other DNA markers (Wright and Bentzen 1995). Particularly with the introduction of automation and fluorometric detection methods, microsatellite assays have become very fast.

1.4.2 Constraints of microsatellite markers

Although microsatellite markers have many positive attributes they are not without constraints. One major disadvantage is that microsatellite isolation requires great effort, time and expense in library construction, screening, sequencing and PCR primer designing (Dunham 2004; Liu and Cordes 2004). In many cases, however, the microsatellite loci along with their primer binding sequences are conserved among closely related species. In these cases, primer designed for one species can be used on others, avoiding the arduous process of their isolation anew.

Rapid processing of microsatellite genotyping requires specialized equipment such as an automated fluorescent sequencer. While the use of automated instruments has tremendously increased the speed of genotyping, it has limited the use of microsatellites in small labs.

Another major constraint associated with microsatellite analysis is genotyping error, which can seriously hamper the inferences in population, parentage and relatedness studies. Genotyping errors may arise from a number of sources (reviewed by Hoffman

and Amos 2005). These include: (i) failure of one allele to amplify due to small quantity or poor quality DNA leading to heterozygous individuals appearing as homozygous; (ii) amplification of non-specific products in the PCR and scoring them as true alleles; (iii) null alleles, resulting from non-amplification of some alleles due to primer site mutation and leading to heterozygote deficiency; and (iv) wrong scoring of allele banding pattern due to stutter band, generated by slippage of Taq polymerase during PCR. Problems associated with null alleles can be solved by re-designing the primers from the flanking sequence (Carvalho and Hauser 1998). Stuttering is especially a problem with dinucleotide loci. This problem can be improved by assaying dinucleotide loci with reduced size (<120 bp) or by selecting for tetranucleotides (O'Reilly *et al.* 1996; Naish and Skibinski 1998).

1.4.3 Polymorphism in microsatellite: mutational mechanisms and models

The polymorphism in microsatellites originates from size variation due to a mutation. Mutation rates at microsatellite loci are generally very high (10^{-2} to 10^{-6} per locus per generation) compared to the rates in a coding locus (10^{-9} to 10^{-10} per locus per generation) (Li *et al.* 2002). Two mechanisms have been proposed to explain the high rates of mutaion in microsatellite loci; one involves slipped-strand mispairing (slippage) during DNA replication and the other involves recombination between DNA strands (reviewed by Hancock 1999; Ellegren 2000; Li *et al.* 2002).

In **replication slippage** the nascent DNA strand dissociates from the template strand during replication of the repeat region and later can reanneal out-of-phase with the template strand. As a result of misalignment, a loop is formed either in the nascent or in the template strand. When replication recommences the new strand becomes longer or shorter than the template strand.

Recombination can potentially alter the lengths of microsatellites in two ways: by unequal crossing-over or by gene conversion. Unequal crossing over takes place when the strands of two homologous chromosomes are misaligned during the chiasma formation. This occurs most easily for long, tandemly repeated sequences where the recombination machinery cannot easily determine the correct register between the two strands. Unequal crossing over gives rise to a deletion in one DNA molecule and insertion in the other. Gene conversion involves unidirectional transfer of information by recombination, probably as a response to DNA damage.

The evidence suggests that replication slippage has the primary role in length mutation in microsatellites (reviewed by Hancock 1999). Key evidence includes:

- Length instability of tandem repeats in *E. coli* remains unaffected by mutants that greatly decrease recombination frequencies.
- Rates of microsatellite mutations are similar in mitotic and meiotic yeast cells, despite much higher recombination rates during meiosis.
- Mutation rates at tandem repeats are much higher than recombination rates.
- Most length mutations at microsatellites represent gains or losses of single repeat units whereas recombination based mutation would be expected to give rise to a wider range of novel mutants.

Various factors can affect the rate of mutations at microsatellite loci; for instance, length of microsatellite, repeat motif, chromosome position, GC content in flanking DNA and age and sex of the individual (Ellegren 2000; Li *et al.* 2002).

While the mutational mechanisms described above explain the molecular basis of mutation, several theoretical models have been proposed to account for the evolution of microsatellites from population data. These include: **Infinite Allele Model (IAM)**,

Stepwise Mutation Model (SMM), **Two Phase Model (TPM)**, and **K-Allele Model (KAM)** (reviewed by Jarne and Lagoda 1996; Estoup *et al.* 2002). IAM postulates that mutation will lead only to new allelic states and can change the length of a microsatellite locus by any number of repeats. Alternatively the SMM predicts that the mutational event leads to a loss or gain of single repeat unit, resulting in the production of some alleles which may already be present in the population. TPM, actually an offshoot of SMM, considers that mutation modifies the current allele size by one unit with approximate probability *P* and by more than one unit with probability 1-*P*. The KAM postulates that there are exactly *K* possible allelic states and any allele has a constant probability of mutating towards any other *K-1* allelic state. Except for the IAM, all of the models take into account the possibility of 'homoplasy' where different copies of a locus are identical in state, although not identical by descent (Balloux and Lugon-Moulin 2002; Estoup *et al.* 2002). It is crucial for the population geneticist to distinguish between the various models of mutations (Jarne and Lagoda 1996), as the correct interpretation of population genetic parameters strongly depends on these models and their associated statistics. Most studies on microsatellite mutation indicate that the TPM is the most realistic mutation model among those defined above (Estoup *et al.* 2002).

1.5 Applications of microsatellite markers in aquaculture

The potential roles of genetic markers, including the microsatellite markers, in management of commercial and rehabilitational aquaculture have been listed in section 1.3. These applications will be discussed in the following sub-sections under three main headings: (i) stock discrimination and population genetics; (ii) parentage and relatedness analysis; (iii) linkage mapping of QTL and selective breeding.

1.5.1 Stock discrimination and population genetics

Prudent management decisions require discrimination of stocks, which allows treating the stocks as separate management units (Moritz 1994). Microsatellites have been frequently used for this purpose, especially where other markers failed to differentiate between populations. Some examples are, Atlantic salmon (Norris *et al.* 1999), cod (*Gadus morhua*) (Ruzzante *et al.* 1999), toothfish (*Dissostichus eleginoides*) (Smith and McVeagh 2000) and gilthead sea bream (*Sparus aurata*) (Alarcón *et al.* 2004), where microsatellites detected more genetic variation compared to allozymes and/or mtDNA markers. In a comparative study between RAPD and microsatellites conducted on common carp (*Cyprinus carpio*) broodstocks, microsatellite analysis revealed more detailed information in various population genetic features like allele frequency, heterozygosity and genetic distance. Particularly, microsatellite analysis showed the presence of some private alleles which the RAPD assay did not show (Bártfai *et al.* 2003).

Microsatellites have been used in numerous cases to study the effect of domestication on genetic composition of fish and shellfish such as in Atlantic salmon (Tessier *et al.* 1997; Crozier 1998; Koljonen *et al.* 2002), turbot, *Scophthalmus maximus* (Coughlan *et al.* 1998), brown trout (Hansen 2002; Was and Wenne 2002), Japanese flounder, *Paralichthys olivaceus* (Sekino *et al.* 2002), common carp (Bártfai *et al.* 2003), black tiger shrimp, *Penaeus monodon* (Xu *et al.* 2001) and abalone, *Haliotis sp.* (Evans *et al.* 2004). Most of these studies reported significant losses of genetic diversity in hatchery reared stocks. Norris *et al.* (1999) found a reduction of 20-48% in allele numbers in farmed Atlantic salmon populations compared to wild ones. According to these workers, founder effects, small population sizes and artificial selection were the likely reasons for such a decrease in allelic number. In a similar study on Japanese flounder,

the average number of alleles per locus for three hatchery stocks was found to be substantially less than that in the corresponding wild populations from where the hatchery stocks originated (5.9-10.7 alleles in hatchery stocks as against 15.3-18.2 alleles in wild populations), even though the sample size analysed for hatchery stocks (100 individuals) was larger than that of each wild population (50-72 individuals) (Sekino *et al.* 2002). Although two of the three hatchery stocks analysed were first generation offspring, significant losses of genetic variation were observed. According to the authors, the reduction in the number of alleles per locus was caused by losses of rare alleles.

In almost all studies, losses of genetic diversity in hatchery populations were observed in terms of microsatellite allele frequency and allele number but not in terms of heterozygosity. O'Connell and Wright (1997) suggested that comparison of allelic diversity is more informative in population studies than estimates of heterozygosity because of the very high level of polymorphism in microsatellite loci. They also suggested that very high levels of variability may confound the usefulness of some microsatellites for comparing the levels of genetic diversity among populations, a similar phenomenon observed with single-locus minisatellite markers. In such cases it has been proposed to perform "allele binning" where alleles within a designated size range are pooled into a composite allele and various population parameters are estimated for allele groups instead of specific alleles (Taylor 1991; O'Reilly and Wright 1995).

Hansen (2002) estimated the long-term impacts of stocking domesticated brown trout (*Salmo trutta*) into rivers containing wild fish by studying two populations. Both the populations were analysed both before and after the stocking event took place. Genetic

variability at nine microsatellite loci was estimated. For analysing the genetic structures of the populations before stocking, historical scale samples collected during the years 1945-1956 were used whereas for post-stocking analysis contemporary samples collected during 1986-2000 were used. In one of the populations the historical and contemporary samples were remarkably similar genetically despite more than a decade of intense stocking. A small genetic introgression (only about 6%) from the domesticated stock was observed compared to an expected contribution of 64%. This indicated poor performance and low fitness of the domesticated trout in the wild. On the contrary, the second population showed a high level of introgression (between 57% and 88% in different samples) from domesticated trout into the wild population. Forced interbreeding of hatchery broodfish with fish collected from wild and their subsequent stocking into the river, combined with low population size of wild trout relative to the number of stocked trout, could explain the high level of introgression.

Detecting hybridization and studying its consequences is another area where microsatellite-based stock discrimination has been useful. In a recent study, Mia *et al.* (2005) detected substantial hybridization between silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) in Bangladesh hatcheries. Three microsatellite markers isolated from silver carp showed species-specific allelic variation between silver and bighead carps and hence were used in this study to detect hybridization. Analysis of 422 hatchery broodfishes, morphologically identified as silver carp showed that 8.3% had bighead allele(s) at one or more loci, while 23.3% of the 236 fish morphologically identified as bighead carp had silver carp alleles. The results suggested that while some of these fish might be F_1 hybrids, others had more complex genotypes, indicating further generations of hybridization or introgression

between the species in hatcheries, with potentially damaging consequences for the integrity of these stocks and their performance in aquaculture.

1.5.2 Parentage and relatedness analysis

Any type of genetic marker can be used for pedigree tracing provided it is sufficiently polymorphic (Gerber *et al.* 2000). Very high levels of polymorphism of microsatellite markers along with the rapid development of computer software packages to handle large amount of genotyping data and development of suitable statistical methods have made these markers an indispensable tool for parentage and relatedness studies (Van de Casteele *et al.* 2001; Liu and Cordes 2004).

The usefulness of microsatellite markers in parentage assignment has been tested on many aquatic species with varying degrees of success. In sea bass (*Dicentrarchus labrax*) parentage could be assigned to 100% of the cases with only two microsatellite loci (Garcia de Leon *et al.* 1998). In African catfish (*Clarias gariepinus*) the rate of success was 85% to 91% with nine microsatellite loci (Volckaert and Hellemans 1999) and in Atlantic salmon the accuracy was over 99.5% with four microsatellite loci (O'Reilly *et al.* 1998). According to Ferguson and Danzmann (1998) the precision of parentage assignment depends not only on the number and variability of microsatellite markers, but also on the number of potential parental pairs from which to choose. The presence of unique alleles for families can be directly useful in parentage determination as was observed by Hara and Sekino (2003) in Japanese flounder, where 100% success rate was obtained in parentage analysis using only four hypervariable microsatellites with many unique alleles. In Kuruma shrimp (*Penaeus japonicus*), simulation based on allele frequencies from 30 dams and 150 putative sires demonstrated that at least five loci would be required to assign progeny to their correct maternal parent with 95%

confidence. Actual assignment however, achieved only 47% success although 6 microsatellites were used. Null alleles and allelic dropout from poor quality DNA contributed to this disparity (Jerry *et al.* 2004).

Marker-based inference about genealogical relationships among individuals has been used in many aquaculture contexts: in estimating heritabilities of commercially important traits, estimating the level of inbreeding and the effective breeding number in populations, minimizing inbreeding in captive populations, and estimating variance in reproductive success among individuals.

Knowledge about heritabilities of important traits can help predict the success of selective breeding. Heritability is a measure of the proportion of phenotypic variance in a population that is predictably passed on to the next generation. Heritability is estimated from the degree of resemblance between relatives (Falconer and Mackay 1996). In heritability studies, it is important to rear the relatives (e.g. half and full-sib families) in the same environment so that the environmental effects on phenotypes does not give a biased estimation of heritability (Fishback *et al.* 2002). If molecular markers are not used, then families need to be reared separately until they are large enough to be physically tagged for communal rearing. However, substantial bias might be introduced during this initial phase of separate rearing. The use of microsatellite markers for pedigree analysis, on the contrary, allows communal rearing from a very early life stage. Microsatellite-assisted pedigree analysis has been applied in heritability studies on a number of important cultured species such as silver crucian carp, *Carassius langsdorfi* (Koedprang *et al.* 2000), rainbow trout (Fishback *et al.* 2002), common carp (Vandeputte *et al.* 2004), Japanese flounder (Shikano 2005) and Nile tilapia, *Oreochromis niloticus* (Charo-Karisa *et al.* 2005).

Sekino *et al.* (2003) used microsatellites to assess the relative contribution of broodfish in a mass spawning experiment of Japanese flounder and observed a highly skewed representation of parents in the next generation. Although 6 males and 12 females were used as broodfish, only 1 male and 6 females contributed to produce offspring. This led to a drastic fall of 80% in effective population size (N_e) after just one generation of breeding. Fiumera *et al.* (2000) estimated N_e in captive-bred haplochromine cichlid *Prognathochromis perrieri* subpopulations using microsatellite markers. Based on the changes in microsatellite allele frequencies, N_e in different subpopulations were estimated to range from 2.5 to 7.7 individuals. These sizes were significantly smaller than actual census sizes (N_{obs}) with corresponding N_e/N_{obs} ratios being in the range of 0.01 to 0.12. Approximately 19% of the initial alleles were lost within the first four generations of captive breeding. To reduce the loss of genetic variation, the workers recommended that offspring production be equalised by periodically removing dominant males, which will encourage reproduction by additional males.

In cases where pedigree information is missing, microsatellite markers have been used to estimate relatedness between individuals. Relatedness (r) is a measure of the fraction of alleles in the genome of two related individuals that are identical by descent (Blouin 2003). Relatedness estimation has been proposed to be a useful tool for (i) estimating heritabilities in nature where partial or no pedigree information is available (Ritland 2000; Blouin 2003); (ii) minimizing inbreeding in population by choosing mates based on r (Lynch and Ritland 1999) and also (iii) to assign pairs to relationship categories (Blouin 2003). The applicability of microsatellites in relatedness analysis have been studied on several species, such as Atlantic salmon (Norris *et al.* 2000), turbot (Borrell *et al.* 2004), rainbow trout (McDonald *et al.* 2004), Japanese flounder (Sekino *et al.* 2004) and Senegal sole, *Solea senegalensis* (Porta *et al.* 2006).

1.5.3 Linkage mapping of QTL and selective breeding

Most of the commercially important quantitative traits of fish, such as growth rate, disease resistance, age of sexual maturation, feed conversion efficiency, etc. are polygenic (Tave 1993; O'Connell and Wright 1997). Traditionally the selection of these traits has involved selecting individuals or families showing better features (e.g. better growth or better disease resistance) for generations. This type of programme requires a large input of time, space and money. The use of highly polymorphic molecular markers can expedite the improvement in selection programmes using a technique referred to as marker-assisted-selection (MAS). MAS may be particularly useful for low-heritability traits and those complicated by dominance effects (Dunham 2004). MAS is expected to increase genetic response by affecting intensity and accuracy of selection (Bentsen and Gjerde 1994; O'Connell and Wright 1997). The first step towards marker-assisted-selection is to create a linkage map of polymorphic loci that adequately covers the whole genome and then to map QTLs based on linkage disequilibrium between alleles at a marker locus and alleles at the linked QTL (Falconer and Mackay 1996).

Type I markers such as isozymes and ESTs are of special importance in linkage mapping due to their inherent advantage of being associated with actual genes. However, type II markers, mainly microsatellite and AFLPs are being used extensively in linkage mapping due to their high polymorphism and genome wide abundance. Microsatellites which are found within genes can offer added advantages for MAS (Dunham 2004; Liu and Cordes 2004). Depending on genome size a large number of markers (maybe several thousands) need to be mapped to place QTLs in tight linkage to the markers (Dunham 2004). A map distance of 20 centimorgan (cM) between adjacent marker loci is the minimum requirement for obtaining a reliable detection of QTLs (Falconer and Mackay 1996; O'Connell and Wright 1997).

Microsatellites have been used solely or in association with other markers in creating linkage maps in many fishes, such as Nile tilapia (Kocher *et al.* 1998), channel catfish (*Ictalurus punctatus*) (Waldbieser *et al.* 2001), rainbow trout (Nichols *et al.* 2003), Atlantic salmon (Gilbey *et al.* 2004), common carp (Sun and Liang 2004) and especially for model species, such as zebra fish (*Danio rerio*) (Shimoda *et al.* 1999) and *Fugu* (Aparicio *et al.* 2002). QTL-linked microsatellite markers for sex determination, growth, feed conversion efficiency, tolerance of bacterial disease, spawning time, embryonic development and cold tolerance have been identified in different species. For example, sex and color QTLs in *Oreochromis* (Kocher *et al.* 2002), sex-determining locus in rainbow trout (Young *et al.* 1998) and sex linked markers in medaka (Matsuda and Nagahama, 2002) have been mapped. Sakamoto *et al.* (1999) have mapped thirteen QTLs of spawning time in rainbow trout using only 54 microsatellite markers. Zimmerman *et al.* (2005) reported identification of three QTLs associated with pyloric caeca number (important for feed conversion efficiency and overall growth) in rainbow trout using F1 haploid hybrid progeny of two clonal lines differing in caeca number. Tanck *et al.* (2001) found 11 microsatellites that were correlated with mass and length in common carp. Ozaki *et al.* (2001) identified several chromosome regions containing putative QTL genes for resistance to infectious pancreatic necrosis (IPN) in rainbow trout using 51 microsatellite markers. Rodriguez *et al.* (2004) used backcrosses of rainbow trout and steelhead (*Oncorhynchus mykiss*) to construct a linkage map and identified a number of AFLPs and microsatellites associated with resistance to infectious hematopoietic necrosis virus (IHNV). QTLs for fitness traits and survival have also been identified in fish. In rainbow trout one microsatellite accounted for 7.5% of the variance in thermal tolerance in unselected populations of rainbow trout (Perry *et al.* 2001). A cold tolerance QTL has been mapped in tilapia hybrid (*Oreochromis*

mossumbicus x *O. aureus*) (Cnaani *et al.* 2003) and in common carp (Sun and Liang 2004).

1.6 Carp culture in Asia with special highlights on South Asia (Bangladesh and India)

Aquaculture in Asia is dominated by different species of carps (fish belonging to Family Cyprinidae). China, India, Bangladesh, Vietnam, Indonesia, and Thailand are the major producers of carps (Table 1.2). The most important species cultured are common carp; Chinese major carps viz. *Hypophthalmichthys molitrix* (silver carp), *Mylopharyngodon piceus* (black carp); *Ctenopharyngodon idella* (grass carp) *and Aristichthys nobilis* (bighead carp); Indian major carps viz. *Catla catla* (catla), *Labeo rohita* (rohu) and *Cirrhinus cirrhosus* (mrigal); and *Barbonymus gonionotus* (silver barb) (Dey *et al.* 2005b). Species which are indigenous to a country are generally the most preferred candidates for aquaculture because of their popularity with local people. However, a number of exotic species have also received widespread popularity among the farmers for their different favourable qualities. In Bangladesh exotic carps now contribute almost 50% of the total aquaculture production. Silver carp is the most popular exotic species, alone contributing about 25% of the total culture production in Bangladesh (FAO 2005). Other instances where introduced species have thrived well in aquaculture systems include common carp in India, Bangladesh, Indonesia and Thailand; grass carp and silver barb in Bangladesh; and Indian major carps in Vietnam (Dey *et al.* 2005b; Reddy 2005; Hussain and Mazid 2005).

Table1.2: Carp production from aquaculture in Asian countries[1]

Country	Carp production (metric tons)	% of total aquaculture production	Year	Major species
China	12 892 221	80.83	2001	*Cyprinus carpio* *Mylopharyngodon piceus* *Ctenopharyngodon idella* *Hypophthalmichthys molitrix* *Aristichthys nobilis* *Carassius carrasius* *Parabramis pekinensis*
India	1 964 287	93.61	2001	*Catla catla* *Labeo rohita* *Cirrhinus cirrhosus*
Bangladesh	530 000	88.55	2001	*Labeo rohita* *Catla catla* *Cirrhinus cirrhosus* *Hypophthalmichthys molitrix* *Cyprinus carpio* *Ctenopharygodon idella*
Vietnam	400 000	29 (approx)	1996	*Hypophthalmichthys molitrix* *Ctenopharyngodon idella* *Aristichthys nobilis* *Labeo rohita* *Cirrhinus cirrhosus* *Cyprinus carpio*
Indonesia	236 363	58.94	2001	*Cyprinus carpio* *Barbonymus gonionotus* *Osteochilus hasseltii*
Thailand	60 199	20.78	2001	*Barbonymus gonionotus* *Cyprinus carpio*

[1] Source: FAO (2003), taken from Dey *et al.* (2005b)

With the development of hypophysation techniques in the 1950s, carp seed for aquaculture are now being supplied almost entirely by hatcheries. In Bangladesh more than 98% of the total spawn production is contributed by hatcheries (Banik and Humayun 1998). With the mushrooming growth of thousands of hatcheries in the region, there is, however, an increasing concern about the genetic quality of hatchery produced seed. The most important causes of genetic deterioration in hatcheries are thought to have been inbreeding and negative selection. Numerous incidences of retarded growth, reduction in reproduction performance, morphological deformities, and increased incidence of disease and mortality of hatchery produced seed have been reported due to these causes (Mair *et al.* 2002; Hussain and Mazid 2005). Indiscriminate hybridisation between species has been identified as another major cause of genetic erosion. High levels of hybridization among three Indian major carps, rohu, catla and mrigal have been reported in some hatcheries in Bangladesh (Simonsen *et al.* 2004) and India (Padhi and Mandal 1994). Systematic studies have shown that most of the Indian major carp hybrids are fertile (Ayyappan *et al.* 2001). This makes it extremely important to prevent indiscriminate hybridization and their unplanned distribution among farmers as this may cause serious damage to the original gene pool through genetic introgression (Padhi and Mandal 1994). Incidental loss of an original gene pool through hybridization has occurred for *Hypophthalmichthys harmandi* in Vietnam, where stocks of this indigenous species are thought to have been entirely replaced by exotic *H. molitrix* (Dan *et al.* 2005). There have also been many reports about hybridization between silver and bighead carps in Bangladesh. This has recently been confirmed by microsatellite analysis (Mia *et al.* 2005).

Every year a large number of carp fry (of both native and exotic species) are released to the floodplains of Bangladesh and India as a means to enhance the capture production

(Reddy 1999; Ahmad *et al.* 1998). Besides, large numbers of farm-reared fishes escape to open water bodies through frequent floods in these regions (Rajts *et al.* 2002). Mixing of poor quality hatchery fish (including hybrid fishes) with wild populations can have serious consequences on the wild gene pool, although no systematic study has so far been conducted on this.

The genetic management situation for non-indigenous species is even more worrying because in most cases the genetic qualities of the founder stocks were unknown (Penman *et al.* 2002). For example some of the stocks of the Chinese carp originally introduced in Bangladesh were secondary or tertiary transfers via other countries such as Nepal and no record was available about their background (Penman *et al.* 2002; Sattar and Das 2002). Most of the current stocks are thought to have experienced bottleneck through very small numbers of founder broodfish and hence are likely to have suffered loss of genetic variation. While for native species it is possible to replace a poor broodstock with fish collected from wild source, in non-indigenous species that is not possible due to logistic and political factors (Penman *et al.* 2002). To improve the condition of exotic species culture, the Department of Fisheries (DOF) of the Bangladesh Government in collaboration with Department for International Development (DFID) imported silver, bighead and grass carp fry directly from the Yangtze River of China in 1994. These new stocks are now being managed and reared separately to maintain pure strains of imported Chinese carps. Attempts are also being made to disseminate fingerlings from these broostocks to different hatcheries (Sattar and Das 2002).

Except in China, genetic research on fish in other Asian countries is still in its infancy. In China some systematic studies have been carried out on population genetics of carps

using molecular markers in different river systems to help make conservation decisions (reviewed by Jian *et al.* 2005; Penman 2005). Hybridization, selective breeding, chromosome manipulation and gene transfer have been applied to improve the cultured stocks in China. In India, initially most of the works were limited to interspecific and intergeneric hybridization with the objectives to produce sterile fish or hybrids with superior culture traits such as faster growth (Ayyappan *et al.* 2001). Chromosome set manipulations have also been undertaken in a number of carp species in countries other than China; the most notable ones being mitotic gynogenesis in rohu, mrigal and silver barb and polyploidy in rohu, catla, common carp and silver barb in India, Bangladesh and Thailand (Penman 2005; Reddy 2005; Hussain and Mazid 2005; Pongthana 2005). The objectives of chromosome manipulation studies were mainly to create monosex or sterile populations to achieve better growth of the culture stocks. In Thailand, an all female population has been created for silver barb (Pongthana 2005).

Successful selective breeding has been reported for some stocks (reviewed by Penman 2005). In India two generations of selective breeding in rohu have resulted in 35-40% faster growth. Broodstocks collected from five different rivers and one farm were used as the base population for this selection programme. In Bangladesh stock improvement through selective breeding is in progress for silver barb, for which a base population was synthesized by crossing among three strains of the species viz. Indonesian, Thai and the existing local stock (of Thai origin). After two generations of selective breeding a total gain of 21.9% has been achieved. In Vietnam, 33% faster growth rate has been achieved in five generations of mass selection in common carp using hybrids of Vietnamese white carp, Hungarian scaled carp and Indonesian yellow carp. However, realized heritability decreased to nearly zero by the sixth generation and hence mass selection approach was replaced by family selection to gain further improvement.

Only a few studies involving molecular markers have been conducted for genetic analysis of carp populations in South-Asia. Isozymes have been used in a few population genetic studies in Indian major carps (Reddy 2005). Isozymes have also been used to identify inadvertent hybridization between Indian and Chinese major carp species in Bangladesh (Simenson *et al.* 2004; Mia *et al.* 2005). Recently, mtDNA RFLP and microsatellites have been used to differentiate between wild and hatchery stocks of carps (reviewed by Reddy 2005). Suitability of RAPD markers for genetic differentiation between four species of Indian major carps has been evaluated (Barman *et al.* 2003).

In view of the current deterioration of the carp genetic resource in South Asia, genetic management techniques are urgently needed to be applied. Particularly important is to develop and take recourse to molecular markers technology to facilitate studies on different aspects of genetic management and improvement of culture stocks.

1.7 Aims and structure of the present thesis

The primary goal of the present Ph.D. research project was to study different aspects of genetic management of carps for aquaculture in South Asia with microsatellite markers and also to generate new markers to facilitate future research. Experiments were conducted on three major culture species: silver carp, common carp and rohu. The project started with characterisation and utilization of microsatellite markers, newly isolated from silver carp by other workers at the Institute of Aquaculture just prior to the commencement of the PhD work. The common carp experiment used a selection of many available markers. And for rohu, new microsatellites were isolated and characterized. Due to limitations in time, the newly isolated rohu microsatellites could not be applied for any specific management related purpose but were only tested for

their polymorphism. These microsatellites will facilitate future research on rohu population genetics and management of hatchery strains while this isolation broadened my experience with different aspects of microsatellites. In brief the present research had the following aims:

1. To characterize newly developed silver carp microsatellites and to cross-amplify them in 10 other carp species.
2. To investigate the usefulness of silver carp microsatellites in parentage, sibship and relatedness analysis.
3. To study the heritability of harvest traits in silver carp using microsatellite markers.
4. To monitor a selective breeding programme in common carp using microsatellite markers.
5. To develop and characterise microsatellite markers in rohu.

The first two aims are addressed together in Chapter 3 and the rest of the aims form single chapters in the thesis.

CHAPTER 2

GENERAL MATERIALS AND METHODS

This chapter describes the basic materials and methods which were of general use throughout the course of the study. Materials and methods specific to certain chapters are described therein.

2.1 DNA extraction

Two methods of DNA extraction were used during the course of the research work depending on what level of purification was needed in the extracted DNA. When high quality purified DNA was essential e.g. for cloning genomic DNA in plasmids, the phenol-chloroform extraction method was employed. For the majority of other purposes, e.g. for fragment analysis, a Chelex extraction protocol was used. The greatest advantage of using the Chelex methods is that it is very fast, although the DNA quality is not very good. DNA extracted by Chelex method is also not suitable for long term storage.

2.1.1 Phenol-Chloroform extraction

Using neutralized phenol or phenol/chloroform is the standard and preferred way to remove proteins from nucleic acid solutions. Tissue samples from different organs were used for DNA extraction. For cloning of the genomic DNA, heart and liver tissues were used whereas for fragment analysis, fin samples were used.

The tissue samples (approximately 50 mg) were placed in individual nucleic acid free 1.5 ml microcentrifuge tubes containing 340 µl of 0.2M EDTA solution (pH 8.0) with 0.5% SDS (sodium lauroylsarcosine, Sigma). Ten µl of 20 mg ml^{-1} proteinase K (ABgene) was added to each tube, mixed briefly and the tube was incubated overnight at 55^{o}C in a hybridization oven (Techne Hybridizer HB-1). During incubation the tubes were tumbled to ensure constant and homogenous mixing. Following this step 10 µl of

20 mg ml^{-1} DNAse free RNAse (ABgene) was added to each tube, which was then shaken vigorously and incubated for 60 min at 37^{o}C in a hybridization oven. About 350-400 µl of buffered phenol (Fisher Scientific) was added to each tube and vortexed for 10 seconds followed by gentle over end turning of the tubes for 15-20 min. About 350-400 µl of chloroform (Fisher Scientific) was added to each tube, shaken vigorously for 10 sec and the tube was then turned for another 15-20 min. The tubes were centrifuged for 5 min at 10,000 g to separate the organic and aqueous phases. About 300 µl of the top aqueous layer was removed to a clean tube, carefully avoiding proteins at the aqueous: organic interface. About 900 µl of chilled 92% ethanol was added to the aqueous solution and mixed by vigorous inversion of the tubes 5-6 times to precipitate the DNA. After allowing the precipitate to stand for 2-3 min most of the ethanol was carefully decanted off. One ml of 70% ethanol was added to wash the DNA pellet and the tubes were then left in a rotator (Stuart Scientific) overnight at room temperature. The ethanol was then carefully decanted off and the DNA was allowed to partially dry for 5-10 min at room temperature before resuspending in 50 µl TE buffer (10 mM Tris.Cl, 1 mM EDTA; pH 8.0). The DNA was stored at -20^{o}C in a freezer.

2.1.2 Chelex extraction

Chelex DNA extraction was carried out in 96-well thermo-resistant PCR plates rather than in individual microcentrifuge tubes. One hundred µl of 10% Chelex (Sigma) solution (made in autoclaved TE buffer, pH 8.0 and containing 0.1% SDS) was added to each well of the PCR plate. Three µl of 10 mg ml^{-1} proteinase K was added to each well. By using a biopsy punch (hole diameter 3 mm) an approximately equal amount of tissue from each fin sample was taken and placed in the Chelex solution. The plate was sealed with adhesive PCR films (ABgene) and incubated for 3 h to overnight at 55^{o}C in a thermocycler to digest the tissue. After digestion the plates were spun for 1 min at 300 g

to bring down all the solution into the bottom of the well. Samples were then incubated at 95^{o}C for 15 min and then again spun down. The extraction was stored at -20^{o}C.

2.2 DNA quantification

Two methods were used for DNA quantification. In the first method quantification was carried out by measuring absorbance or optical density of the DNA solution in a spectrophotometer (Jenway Ltd.) at 260 nm. DNA concentration was calculated by the following formula:

DNA Concentration (ng/µl) = Optical density x dilution x 50

Quantification by the second method involved running 3 µl of a DNA sample in a 1.2% agarose gel for 5 min at 5 V/cm (see section 2.3) and comparing the intensity of the EtBr-stained bands against those of different amounts of marker φ174 RF DNA *Hae* III (ABgene) viz, 50 ng, 100 ng, 150 ng, 200 ng, 250 ng run on the same gel.

The first technique can only be used on purified DNA, such as DNA extracted by phenol-chloroform, whereas the second method is suitable for almost all types of DNA. Upon quantification the DNA solutions were brought to the desired concentration level by addition of TE buffer (pH 8.0).

2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was routinely used for checking the success of DNA extraction, PCR amplification etc. For routine use 1-1.2% agarose gel was used with 0.5X TAE buffer (50X stock solution contained 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml of 0.5 M EDTA (pH 0.8) in 1 litre solution) or 1X Sodium boric acid buffer (20X stock solution was prepared by dissolving 0.8 g NaOH, pH adjusted to 8.5

by boric acid). Sodium boric acid buffer was preferred over TAE because less heat was generated during electrophoresis, allowing a much higher voltage (5-35 V/cm) to be used compared to TAE (5-10V/cm) (Brody and Kern, 2004). This drastically reduced gel run times. Gels were stained with Ethidium bromide (0.5 µg/ml; Sigma) to visualize the DNA bands under a UV transilluminator.

2.4 Optimization of Polymerase Chain Reaction (PCR)

Optimization of the annealing temperature (T_a) is generally the most crucial step in optimization of PCR. T_a was optimized by testing amplification in a gradient of temperatures in the range of 10^oC below the estimated melting temperature (T_m) in a Biometra TGradient Thermocycler (Thistle Scientific). The temperature giving the best amplification was chosen for subsequent PCRs.

Initially all PCRs were attempted with the following reagent composition: 1x PCR buffer IV (75 mM Tris-HCl, pH 8.8; 2 mM $(NH_4)SO_4$, 0.01% (v/v) Tween 20®); dNTPs 150 µM each, 1.5 mM $MgCl_2$, forward and reverse primers 0.15 pmol/µl each, 0.2 U Taq DNA polymerase, 70 ng DNA and water to make the total volume 15 µl. The buffer, $MgCl_2$, dNTP set and DNA polymerase were procured from ABgene and the primers from MWG Biotech. Depending upon the success of amplification some adjustments in the concentrations of the reagents were carried out as follows: $MgCl_2$ between 1.5 to 2.5 mM, dNTPs between 150 µM to 250 µM, primers between 0.15 pmol/µl and 0.2 pmol/µl and Taq DNA polymerase between 0.2 to 0.5 U.

2.5 Optimization of multiplex PCR

Multiplex PCR allows simultaneous amplification of several loci in the same PCR reaction. Multiplex PCR was optimized for fragment analysis using fluorescent primers.

A preliminary check for compatibility of the primers for multiplexing was carried out using the program 'PrimerSelect' from Lasergene: Expert Sequence Analysis Software, DNAstar Inc., Ver.5. This program looked for primer pair dimers and reported free energy value (ΔG) for each pair. Groups of microsatellite loci to be multiplexed were chosen based on the lowest ΔG value, similar T_a, non-overlapping size ranges or if sizes overlapped then choosing primers which were tagged with different fluorescent dyes.

Optimization of the multiplex PCR was carried out using a step-by-step protocol described by Fishback *et al.* (1999). PCR was performed in a volume of 15 µl. Only the concentrations of the primers and the Taq DNA polymerase were varied to get the optimum levels while the concentrations of other components were kept fixed as follows: 280 µM of each dNTP, 2X PCR buffer II (20 mM Tris-HCl, pH 8.3; 100 mM KCl) (ABgene), 2 mM $MgCl_2$ and 100 ng of genomic DNA. In the first step of optimization all the primers were included in equal concentration of 0.05 pmol/µl. Depending on the intensity of amplification, individual primer pair concentrations were increased or decreased to give similar levels of amplification for all the loci.

The Taq polymerase concentrations were varied between 0.5 U to 2.0 U (increased by 0.5 units each time) to choose the best concentration to get a consistent 'plus A' modification of the alleles and to eliminate the production of spurious artifact bands.

2.6 Fragment analysis

The fragment analysis was performed on denaturing polyacrylamide gels and the gel image was analysed using semi-automated sequencers. For experiments involving silver carp and common carp an ABI^{TM} PRISM 377 DNA sequencer (Applied Biosystem, Perkin-Elmer) was used while for rohu a Beckman-Coulter CEQ^{TM} 8000 Genetic Analysis System was used, as the former instrument was replaced by the latter one in

our laboratory. Both the sequencers are automated instruments designed for analyzing fluorescently-labelled DNA fragments by gel electrophoresis. Although the electrophoresis and analysis principles are essentially the same for both the apparatus, still they differ in some important features. Brief descriptions for both are given below.

2.6.1 ABI™ PRISM 377 DNA sequencer

2.6.1.1 Theory of operation

The ABI PRISM 377 DNA sequencer uses an argon-ion laser to detect fluorescence-labelled fragments of DNA as they pass through the laser read region during electrophoresis. The laser excites the fluorescent dyes attached to the fragments and they emit light at a specific wavelength for each dye. The collected light is separated according to wavelength range by a spectrograph onto a cooled, charged coupled device (CCD) camera, so that different types of fluorescent emissions can be detected with one pass of the laser. The data collection software collects the light intensities from the CCD and stores them into a gel file as digital signals. At the end of data collection, the GeneScan Analysis software is used manually or automatically to process, analyse and translate the collected data into fragment sizing information.

2.6.1.2 Preparation and pouring of gel

The fragment analysis was performed on 6% denaturing polyacrylamide gels electrophoresed in the ABI 377 DNA sequencer. The gel was prepared by dissolving 18 g of urea (Bio-Rad Laboratories) in 5 ml of Long Ranger® gel solution (Acrylamide, 50% stock solution, Cambrex), 5 ml of 10x TBE (108 g Tris, 55 g boric acid, 8.3 g EDTA in 1 litre solution) and 26 ml of distilled water. To remove charged particles from the gel solution 0.5 g of resin beads (Sigma) was added and the mixture was stirred on magnetic stirrer for about 20 min. The solution was then filtered with

cellulose nitrate membrane filters (Whatman) with 0.2 µM mesh size and degassed for about 5 minutes by a vacuum pump. After degassing, 250 µl of 0.1% APS (Ammonium Persulphate, Amresco) and 35 µl of TEMED (N,N,N$^{/}$,N$^{/}$-tetramethylethylenediamine, Sigma-Aldrich) were promptly added into the gel solution and mixed by gentle swirling to avoid introduction of air. Since gel solution cannot be left standing after APS and TEMED have been added, arrangements for pouring the gel were made before the preparation of the gel solution. Two 42 cm long glass plates were washed, dried and set parallel on each other with the bottom ends of the plates leveled together so that the gel can be poured in between the plates. To maintain the thickness of the gel two 0.2 mm thick spacers were placed between the long edges of the plates. The plates were mounted in the gel cassette and the gel injection device was attached to it according to the ABI User's Manual. Once the gel solution was prepared it was immediately injected in between the plates with the aid of a 50 ml syringe.

The gel was left on the bench for about 2 hours to polymerize. The gel was then mounted in the ABI sequencer according to the instructions in the ABI user's manual. The upper and lower buffer tanks were put in place and filled with 1x TBE buffer. The gel was then ready for use.

2.6.1.3 Preparation of samples

To prepare samples for fragment analysis, PCR was conducted using primers labelled with one of the three fluorescent dyes: FAM (6-carboxyfluorescein), TET (6-tetrachlorofluorescein), HEX (6-hexachlorofluorescein). For each locus amplified, only one of the two primers was labeled with dye. Before loading into the gel the PCR samples were diluted with distilled water up to 10 times based on the intensity of amplification assessed on agarose gels and also the fluorescent dye attached. Generally

PCR products with blue dye needed most dilution, followed by green and yellow. About 0.5 µl of diluted PCR product was mixed with 0.5 µl of GeneScan™-350 TAMRA™ size standard (Applied Biosystems, UK) and 1.5 µl of loading solution. The loading solution was prepared fresh before starting the whole process by mixing 5 parts of deionised formamide (pH>7.0) with 1 part of the mixture of EDTA (25 mM) with blue dextran (50 mg/ml). Samples were denatured for 5 min at 95^{o}C and immediately cooled on ice until loading.

2.6.1.4 Loading of samples and electrophoresis

The ABI User's Manual was followed for operation of the sequencer instrument and starting the GeneScan® 3.1.2 programme for running the gel. Pre-electrophoresis steps included a plate check to ensure the plates were clean, performing a pre-run to bring the gel to the right temperature and preparing the sample sheet. After the pre-run the samples were loaded into the gel by a membrane comb. The elecrophoresis run was performed under the following conditions: voltage 1.00 kV, temperature 51^{o}C, and laser power 40 mW. The upper limits for electrophoresis current and power were kept at 35 mA and 50 W respectively. The run time for electrophoresis was 2.0 hours.

2.6.1.5 Processing and analysing the raw data

The raw data stored in the gel file was processed and analysed using the Genescan® Analysis software ver.2.1. It consisted of the following steps: 1) installing the right matrix (created according to the Genescan User's Manual) into the gel image to help multicomponenting, 2) tracking lanes, 3) extracting lanes, 4) defining size standard, 5) analyzing all the samples against the size standard using a predefined analysis parameter. Picking up the right alleles, labeling, and their categorization were

performed using the programme ABI Prism Genotyper Ver.2.0. The analysis of the alleles involved the following steps:

Labelling and filtering peaks

These tasks were performed either manually by selecting the peaks of choice or automated by using the commands "label peaks" and "filter peaks" in the Genotyper programme. During filtration, criteria were set such that any peak less than 32% of the height of the highest peak in a locus's size range or a peak preceeded by a higher peak within 1.6 bp or followed by a higher peak within 3 bp were removed. In most cases this would successfully remove labels from unwanted peaks due to stutter, 'A' nucleotide addition and nonspecific amplification. However, after each automated peak labelling they were checked individually and corrected manually if needed.

Binning and categorization

In this step, the size range of a marker locus was specified. Bin size was specified between 0.1 and 0.3 according to the size range of alleles. The bin size defines an interval within which genotyper calculates a count and frequency of each occurrence of labeled peak data. Lower bin size gives better precision but if set too low (when the allele size range is quite large) then large number of bins will be created making histogram reading difficult. New categories for alleles were created by taking peaks that formed a distinct cluster and also differed by not more than 2 base pairs. Macros and templates were created to ease the performance of the same tasks on several gels.

2.6.2 Beckman-Coulter CEQ^{TM} 8000T

2.6.2.1 Theory of operation

The CEQ 8000 uses the capillary electrophoresis mechanism to separate the DNA fragments instead of the slab gel used in the ABI 377. Capillary electrophoresis occurs

when an electric field is applied to an electrolyte solution within a capillary, causing ions, e.g. DNA, to migrate. DNA fragments are separated as they migrate by a size exclusion sieving effect. The dye labelled DNA fragments are detected by fluorescence and in turn are rendered into fragment sizes. CEQ 8000 uses polyacrylamide gels as the electrolyte solution to provide the sieving medium for the separation. Ready-to-use gel cartridges from Beckman-Coulter were used, obviating the steps of gel preparation. Generally three Beckman Wellred fluorescent dyes viz. Dye2PA (black), Dye3PA (green) and Dye4PA (blue) were used to label the primers whereas Dye1 (red) was used for the size standard. In the present study, rather than ordering individual primers with fluorescent labels, M13 tailed primers and M13 primers labeled with above mentioned dyes were used to add fluorescence to the PCR product. This technique is described in detail in section 2.7.

2.6.2.2 Preparation of samples

After carrying out PCR and necessary dilution of the PCR product (as described in section 2.6.1.3) samples for fragment analysis were prepared by mixing 0.5 µl of PCR product with 0.25 µl of size standard (SS-400) and 30 µl of SLS (Sample Loading Solution, Beckman Coulter Inc.). The samples were prepared in a 96 well sample tray and on the top of each sample 1 drop of mineral oil was added to prevent evaporation. A 96-well buffer tray was prepared by filling two thirds of each well with separation buffer (Beckman Coulter Inc.)

2.6.2.3 Running the samples

Prior to running the samples, a sample sheet was prepared. The capillary set, gel cartridge and sample and buffer plates were installed in the sequencer according to the instructions in the CEQ 8000 series Fragment Analysis Training Guide (2004).

2.6.2.4 Analysing raw data

Raw data were analysed using pre-defined analysis parameters specifying the correct size standard, dye mobility calibration and the analysis method. For the present study the parameters were set as follows: SS-400 as size standard, mobility calibration for phosphoramadite dye and Cubic method of analysis. Analysing raw data labeled all observed peaks. Unwanted peaks were removed either by manual exclusion or by automatic exclusion (filtering of result data).

2.6.2.5 Creating locus tag

Similar to categorization in the ABI softwares, here also locus tags were created for individual markers by specifying the size range of alleles, maximum bin width (generally kept at 1.9) and repeat unit length for the microsatellite. Manual correction was done as necessary to call the correct peaks. Once a new locus tag had been created, the samples were reanalyzed to add the locus tag in the study.

2.7 Use of M13 tailed primer for CEQ

Ordering a large number of fluorescent primers is very expensive. A much less expensive way is to order only one universal primer such as M13 with fluorescence (5′ Dye-CACGACGTTGTAAAACGAC) and to use it for adding fluorescence to the PCR products of any microsatellite locus. To employ this method (described by Beckman-Coulter 2004) one of the primers (e.g. forward primer) of a locus also needs to have an M13 sequence (same sequence as above) added to its 5′ end. This is then called an M13 tailed primer.

In this method, PCR is performed with three primers: M13-tailed forward primer (0.3 pmoles), a non-tailed reverse primer (3 pmoles) and a fluorescently labelled M13 primer

(3 pmoles). Initially, amplification is primed from the M13-tailed primer and the reverse primer on the original template DNA. DNA fragments generated in the first round of synthesis become the template for further amplification with either the forward or reverse primers. By the second round of amplification, the product being amplified will have the M13–tail sequence incorporated into the PCR product. M13 labelled primer, being present in 10-fold excess, will take the place of M13-tailed forward primer by the third round of amplification to create products containing the fluorescence dye that can be detected by the CEQ sequencer. Figure 2.1 shows the process in detail. This technique was used for screening the large number of microsatellites isolated from rohu (Chapter 6).

2.8 Microsatellite marker profiling

Profiling of individual microsatellite markers is important to assess their usefulness and informativeness. The important features in marker profiling include: number of alleles (n), effective number of alleles (A_e), allele frequency, polymorphic information content (PIC), expected heterozygosity (H_e) and observed heterozygosity (H_o), null allele prediction and testing Hardy-Weinberg equilibrium.

2.8.1 Effective number of alleles (A_e)

Effective number of alleles of a locus is calculated based on the allele frequency distribution using the following formula by Ferguson (1980):

$A_e = \frac{1}{\sum q_i^2}$ where q_i is the frequency of the i^{th} allele at a locus. The effective number of alleles (A_e) is an important parameter of the informativeness of a locus.

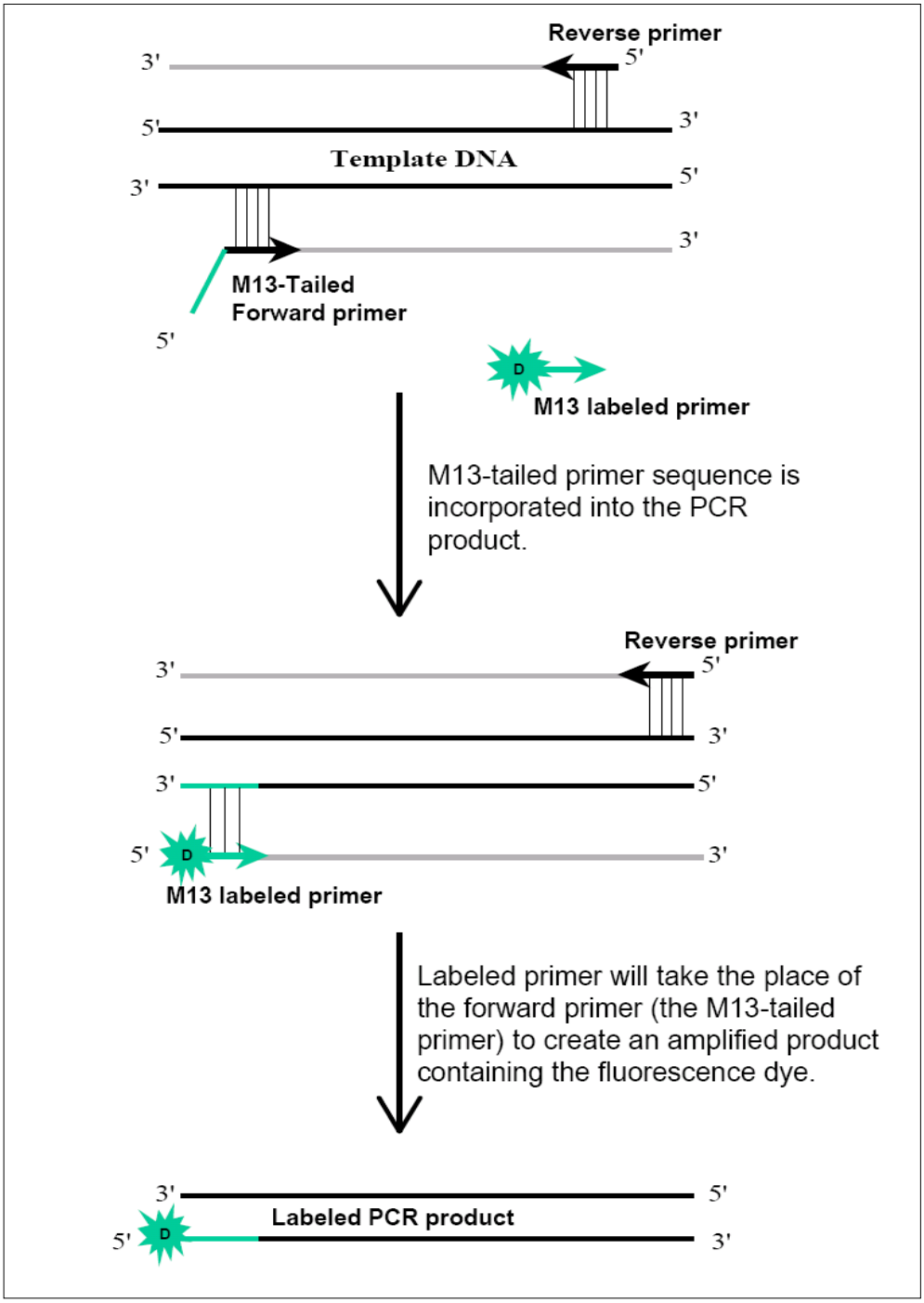

Figure 2.1: Labelling PCR product with fluorescent dye using an M13-tailed primer (Source: Beckman-Coulter, available at www.mafes.msstate.edu/biotech/TailedPrimers.pdf)

2.8.2 Heterozygosity

The expected heterozygosity (H_e) and observed heterozygosity (H_o) were estimated using the programme CERVUS (Marshall *et al.* 1998) or GENEPOP (Raymond and Rousset 1995). The expected heterozygosity of a locus is defined as the estimated fraction of all individuals that should be heterozygous for that locus. H_e is estimated based on the allele frequency distribution as follows: s

$$H_e = 1 - \sum_{i=1}^{n} p_i^2$$ where, P_i is the frequency of the i^{th} of n alleles.

Expected heterozygosity over several loci can be calculated as:

$$H_e = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{n} p_{li}^2$$ where, the first summation is for the l^{th} of m loci.

Expected heterozygosity is also an important measure of microsatellite informativeness and is maximal when the allele frequency distribution is uniform. Observed heterozygosity (H_o) is defined as the observed fraction of the heterozygous individuals for a locus.

2.8.3 Polymorphic Information content (PIC)

The PIC was first developed by Botstein *et al.* (1980) for the case of rare dominant disease. It is now frequently used for measuring the informativeness of a genetic marker. The PIC can be defined as the probability that one could identify which homologue of a given parent was transmitted to a given offspring, the other parent being genotyped as well. The PIC value of an *n*-allele locus can be calculated as:

$$PIC = 1 - \sum_{i} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2$$ where, p_i is the population frequency of the i^{th} allele and n is the number of alleles. PIC is always smaller than H_e. PIC was calculated by the programme CERVUS.

2.8.4 Hardy-Weinberg Equilibrium (HWE)

HWE describes the expected frequencies of genotypes in a population under random mating. The programme CERVUS can analyse the deviation of genotypes at a microsatellie locus from HWE using a chi-square goodness-of-fit, which compares the observed genotype frequencies with the expected frequencies. Deviation from HWE may arise from population substructure (e.g. family groups, mixed subpopulation, hybridization between subspecies), selection acting on linked loci, biases toward typing a particular genotype or null alleles segregating in the population. Population substructure is likely to lead to deviations from HWE at all loci, whereas other causes are mostly locus-specific.

2.8.5 Null allele frequency

The programme CERVUS can estimate the frequency of possible null alleles segregating at each locus using an iterative algorithm based on the difference between the observed and expected frequency of homozygotes. When there is an excess of homozygotes over that which would be expected under Hardy-Weinberg equilibrium, that produces a large positive estimate of null allele frequency. However, a large positive estimate is not a certain proof of presence of null allele because the excess homozygosity in a locus may be created by other causes such as presence of sex-linked locus, bias against typing individuals as heterozygotes, selection acting on a locus or due to mixing of genetically distinct groups in the population under study (the same causes that leads to perturbation in HWE). A positive estimate of null frequency, however, can serve as a warning and allows reconsidering the use of the particular locus with high null allele prediction (above 0.05).

2.9 Parentage assignment

The Family Assignment Programme (FAP) version 3.1a, created by Dr. J. B. Taggart in the Institute of Aquaculture, University of Stirling, was used for parentage assignment from the genotyping data of potential parents and offspring. FAP allowed the performance of two related tasks using the exclusion principle. The first was prediction analysis about the resolving power of specific parental genotypic data sets for unambiguously discriminating among families or groups of families. This was achieved by a complete enumeration of all possible genotypic combinations, using an efficient comparison algorithm. The Second task was actual assignment to family of progeny from known parental and progeny genotypic data. Both the tasks assumed that all the individuals were the progeny of known parental combinations for which full genotypic data was available. It was also assumed that the nuclear loci used in the analysis were independently inherited in simple Mendelian fashion.

CHAPTER 3

CHARACTERIZATION OF SILVER CARP (*Hypophthalmichthys molitrix*) MICROSATELLITES AND THEIR UTILITY IN KINSHIP ANALYSIS

3.1 Introduction

Just prior to the commencement of the present PhD work, a number of microsatellite markers were isolated from silver carp genomic libraries by Dr. John Taggart of the Institute of Aquaculture. The genomic libraries were developed both by the conventional method (without enrichment for microsatellite repeats) and by microsatellite enrichment. The enrichment was undertaken for a variety of tetra and trinucleotide microsatellite motifs-$(GATA)_n$, $(GACA)_n$, $(GGAT)_n$, $(GGA)_n$ and $(AGC)_n$. Although the enrichment probes identified small stretches of tetra and trinucleotide arrays from all the sequences, most of them contained dinucleotide tandem repeats. Primer pairs could be designed from the unique flanking sequences of forty microsatellite repeats. For the research work in the present thesis, some of these markers were characterized for their inheritance pattern and polymorphism. A number of markers were then chosen based on their levels of polymorphism, informativeness and ease of amplification and were incorporated into multiplex PCRs for use in pedigree analysis.

3.1.1 Microsatellite based kinship analysis

Marker-assisted kinship analysis can be divided into three categories.

(i) Assignment of parent or parent pairs to individuals when the potential parents are known.

(ii) Reconstruction of relationship categories when potential parents are unknown but the potential relationship categories in a group of individuals are known.

(iii) Estimation of relatedness when there is no *a priori* knowledge about the parentage or relationship categories.

3.1.1.1 Parentage analysis

Different methods of parentage assignment have been devised to serve different situations and many computer programmes have been developed to implement these methods. Jones and Arden (2003) give a comprehensive review on the available methods of parentage analysis in natural populations. The earliest and conceptually the simplest technique of parentage analysis is ***Exclusion.*** This method is based on strict Mendelian rules of inheritance and uses incompatibilities between the genotypes of candidate parents and offspring to reject particular parent-offspring hypothesis. Since all but one parent (when the other parent is known) or parent pair from a complete set of many putative parents (or pairs) are excluded, this technique theoretically, offers the most precise parentage assignment. However, in practical situations many factors such genotyping error, null alleles and mutations can lead to false exclusion, reducing the applicability of this method. These problems become more acute as more genotype data are brought into the analysis, because the assay of additional loci (or additional individuals) increases the likelihood that a dataset will contain errors or mutations. Under strict exclusion, a single mismatch is enough to exclude a candidate parent. However, most exclusion-based computer programmes allow users to specify the number of mismatches that should be tolerated before a parent pair is excluded, making the method more robust. This technique is most powerful when few candidate parents are involved and highly polymorphic markers are available.

The ***Likelihood method*** has been developed for cases where the exclusion method fails to resolve paternity. Likelihood approaches involve selecting the most likely parents based on the genotypes of the candidate parents and the allele frequency distribution of the marker loci. Contrary to strict exclusion methods, likelihood-based allocation methods usually allow for some degree of genotyping error. Using likelihood method

parentage can be assigned either by the ***Categorical allocation approach*** or by the ***Fractional allocation approach.*** Categorical allocation calculates a logarithm of the likelihood ratio (LOD score) by determining the likelihood of an individual (or pair of individuals) being the parent(s) of a given offspring divided by the likelihood of these individuals being unrelated. A LOD score is determined for all candidate parents (or pairs) and the offspring are assigned to the parent (or pair) with the highest LOD score. Simulation procedures can be applied to determine the significance of results (Marshall *et al.* 1998). When all parent-offspring relationships show zero likelihood, offspring are left unassigned. Parentage remains ambiguous when multiple individuals show the same likelihood.

While the categorical technique assigns the entire individual to a particular male, the fractional technique splits an offspring (some fraction between 0 and 1) among all compatible males (or females). The portion of an offspring allocated to a particular candidate parent is proportional to its likelihood of being the parent of the offspring relative to all other candidate parents. The likelihood of a parent (or parent pair) is calculated essentially the same way as the categorical method. Categorical approaches have been more popular compared to fractional approaches as the analysis by the former method represents biological truth. On the contrary, the fractional technique is considered to have better statistical properties for the evaluation of certain hypotheses even though the technique does not represent exact biological truth. This approach can be useful to provide more precise analysis of the proportion of offspring parented by each of the adults in a population (Neff *et al.* 2001), of the relative reproductive success of males (Nielsen *et al.* 2001; Signorovitch & Nielsen 2002) and can be useful for incorporating prior information about the biology and behaviour of the species into the analysis (Neff *et al.* 2001).

While the exclusion and likelihood methods deal with the assignment of offspring to parent or parental pairs, the ***parental reconstruction*** method uses a totally different approach for parentage analysis. In this method, the genotype of one parent for a group of sibs is reconstructed when the other parent is certainly known. This method was developed to study family structure in natural populations for species where a known parent can be collected with a large group of its progeny (either half- or full-sibs). Reconstruction of the genotype of the other parent from the genotypic information of the known parent and the progeny then allows detecting possible multiple mating (Jones 2001).

A number of software package are now available for parentage analysis. The major ones include: PROBMAX (Danzmann 1997) and FAP (Dr. J. B. Taggart, Institute of Aquaculture, unpublished) for exclusion approach; CERVUS (Marshall *et al.* 1998), FAMOZ (Gerber *et al.* 2000) and PAPA (Duchesene *et al.* 2002) for categorical allocation; PATRI (Signorovitch and Nielsen 2002) for fractional allocation and GERUD (Jones 2001) for parental reconstruction.

3.1.1.2 Reconstruction of relationship categories

Parental samples can be missing under many circumstances. Methods have been developed to reconstruct relationship categories when parental information is absent but there are *a priori* ideas about the possible relationship categories in the sampled individuals. Two approaches have been applied for pedigree reconstruction. One involves pairwise analysis for inferring the most likely relationship between individuals in a dyad (pair). The log likelihood ratio (LOD score) of two competing categories is used to choose the most likely relationship between the dyad (Mousseau *et al.* 1998; Blouin 2003). The advantage of pairwise approaches is that they are simple to

implement but the drawback is that valuable information is lost while considering a pair in isolation of all other individuals in the group (Thomas and Hill 2000). All individuals in a sample may provide direct and indirect information concerning the relationship of a dyad, especially those individuals which are closely related to the dyad. A pairwise relationship approach is sufficient in some instances, for example, when they are used to avoid mating between relatives. However, when it is needed to partition the sampled individuals in distinct genetic groups, for instance, for estimation of genetic parameters, pairwise estimation is not an efficient method. To overcome these problems the group-likelihood approach has been devised (Painter 1997; Thomas and Hill 2000; Smith *et al.* 2001; Konovalov *et al.* 2004). This approach considers all individuals in the entire sample and partitions them simultaneously into distinctive genetic groups.

3.1.1.3 Relatedness analysis

Estimation of pairwise relatedness estimation is a useful tool for kinship analysis when neither parentage information nor information about possible relationships between individuals are known (Blouin *et al.* 1996; Norris *et al.* 2000). Relatedness is a continuous measure of overall identity by descent. Alleles are considered identical by descent when they are identical copies of the same ancestral allele within a particular reference population. What proportion of the alleles between two individuals are identical by descent depends on the type of relationship between the pair. Parent-offspring and full-sibs share 50% of the alleles identical by descent, on average (first degree relatives), grandparents-grandoffspring, half-sibs and avunculars (uncles or aunts with nephews and nieces) share, on average, 25% alleles (second degree relatives) and first cousins and great-grandparent with great-grandoffspring share, on average, 12.5% alleles (third degree relatives). The probabilities of a dyad sharing 0, 1 or 2 alleles that are identical by descent at any locus are expressed by IBD coefficients k_0, k_1, k_2

respectively. The IBD coefficients are combined to obtain the relatedness, *r* (also called coefficient of relatedness) between two individuals and can be expressed as: $r = k_1/2 + k_2$ (Blouin 2003). IBD coefficients and relatedness for common relationship categories are listed in Table 3.1.

Table 3.1 Identity by descent coefficients (k_0, k_1, k_2) and relatedness *r*, for some common relationship categories

Relationship categories	k_0	k_1	k_2	*r*
Monozygotic twins or self	0	0	1	1
Parent-offspring	0	1	0	0.50
Full-sibs	0.25	0.50	0.25	0.50
2^o (e.g. half-sibs, avuncular)	0.50	0.50	0	0.25
3^o (e.g. first cousins)	0.75	0.25	0	0.125
Unrelated	1	0	0	0

A number of Methods-of-Moment (MOM) estimators have been developed to estimate relatedness from genetic marker data. The important ones are: (1) similarity index of Li *et al.* (1993), (2) a regression-based estimator of Queller and Goodnight (1989), (3) a correlation-based estimator of Ritland (1996) and (4) a regression-based estimator of Lynch and Ritland (1999). Unbiased estimation of relatedness by all the above estimators requires information on population allele frequencies. In practice, however, often actual population allele frequency is unknown and is estimated from samples. Estimated allele frequency can be prone to sampling error, especially for highly polymorphic markers like microsatellites. Wang (2002) developed a new estimator (which is originally a modification of the similarity index of Li *et al.* 1993) which can be corrected for the allele frequencies estimated from samples. Maximum likelihood based relatedness estimation has also been proposed (Thompson 1975; Milligan 2003) although the MOM estimators are still more popular.

The relative performance of the different relatedness estimators and their many versions has been studied by a number of workers using simulated and real life genotype data (Lynch and Ritland 1999; Van de Casteele *et al.* 2001; Wang 2002; Milligan 2003). These studies show that the absolute and the relative performances (measured by sampling variance) of the estimators may vary with a number of factors such as the shape of allele frequency distributions, the number of alleles per locus, number of loci used and the population composition. Considering all factors the estimator of Wang (2002) has the most desirable properties, which include low sensitivity to the sampling error that results from the estimation of population allele frequency, consistency over difference in sample size, robustness of unknown relatives being included in estimating allele frequencies and asymptotic decrease of sampling variance to the minimum with an increased number of loci and alleles (Wang 2002; Blouin 2003). Queller and Goodnights' (1989) estimator also performs well and so far has been the most widely used estimator. One undesirable property of this estimator is that it is undefined for heterozygote at a biallelic locus.

Relatedness estimations are typically characterised by very large variances. There are two major sources of variance. First, the observed frequencies of IBD in a pair of relatives might vary across loci although the expected probability of IBD is the same over all loci for a given relationship. Second, alleles may be identical in state but not identical by descent (Wang 2002). According to Blouin (2003) about 30-40 microsatellite loci are needed to obtain even moderate confidence around a single pairwise estimate (e.g. standard deviations of 0.1).

A number of computer packages are now available to calculate relatedness based on different estimators such as KINSHIP (Goodnight and Queller 1999); RELATEDNESS

(Queller and Goodnight 1989), SPAGEDI (Hardy and Vekemans 2002), Mer (Wang, 2002) and DELIRIOUS (Stone and Björklund 2001).

3.1.2 Cross species amplification of microsatellite markers

Isolation of species-specific microsatellites is a tedious and expensive process and hence often becomes a limiting factor for their use. Many studies however, have demonstrated that microsatellite repeats as well as their priming sites are often conserved between closely related taxa, allowing primers designed for one species to be used on another species. In general, the chance of a successful cross-species (heterologous) amplification of any DNA sequence is inversely related to the evolutionary distance between the two species (Primmer *et al.* 1996). A number of studies, however, have demonstrated microsatellite conservation over large evolutionary distances. For instance, 86.2% of the microsatellites from one species of python amplified in other species which last shared a common ancestor at least 40 million years ago (Jordan *et al.* 2002); in turtles homologous microsatellite loci have been conserved for about 300 million years (FitzSimmons *et al*. 1995) and in fish for about 470 million years (Rico *et al.* 1996). Based on their studies on pythons, Jordan *et al.* (2002) suggested that there are selective constraints on the nucleotide substitution process in microsatellite flanking regions. Apart from the evolutionary distance, some other factors may also affect the success in cross-species amplification of microsatellite loci. These include length of repeat, heterozygosity in original species, GC richness of the flanking regions, repeat motifs, and product length, although the significance of these factors may vary in different organisms (Ezenwa *et al.* 1998).

In general, microsatellites have been found to be highly conserved among species of fish (Dunham 2004). Based on fish gene maps generated to date, Liu *et al.* (2003)

suggested that microsatellite flanking sequences of fish may evolve at a slower rate than those of mammals. Thus the ability to use microsatellite primers developed from one species of fish to amplify homologous sequence in other species can greatly reduce the cost and labour of newly isolating microsatellites. Apart from this, conservation of microsatellites across a broad range of taxa in fish can potentially be used for other purposes such as comparative gene mapping, cloning of genes and for evolutionary studies (Dunham 2004).

3.1.3 Objectives of the study

This chapter focuses on the following objectives:

1. Characterization of silver carp specific microsatellite markers.
2. Optimization of multiplex PCR with groups of polymorphic markers.
3. Assessing the usefulness of silver carp microsatellites for parentage analysis.
4. Evaluating the utility of the markers in sibship analysis without parental information.
5. Evaluating the usefulness of the markers for relatedness estimation.
6. Cross amplification of silver carp microsatellites in 10 other cyprinid species.

3.2 Materials and methods

3.2.1 Characterization of silver carp microsatellites

The present work was started with characterization of 20 silver carp microsatellite markers which were selected based on some preliminary work by others in the Institute of Aquaculture. Initially the loci were checked for their inheritance pattern (whether they follow Mendelian transmission or not), polymorphism and reproducibility of PCR amplification by genotyping them on six full-sib families (15 individuals tested per family) with known parents. These families were created by single pair crossing between 6 male and 6 female silver carp broodfish. The breeding was carried out at the NFRDMP (North-West Fisheries Resource Development and Management Project) hatchery at Parbatipur, Bangladesh. For DNA extraction and microsatellite genotyping, fin clips from parents and one day old offspring (about 100 hatchlings) from each family were preserved in 95% ethanol in labelled vials. The parental DNA was extracted by the phenol-chloroform protocol and offspring DNA by Chelex method using whole individual hatchling. The PCR amplification was performed with the following thermocycling conditions:

95°C	2 min 30 sec	
94°C	50 sec	30 cycles
T_a	50 sec	
72°C	1 min	
72°C	35 min	
10°C	hold	

The final extension time at 72°C was kept at 35 min because some preliminary optimization work on PCR showed that lower extension times allowed variable A^+ addition at the end of PCR products, creating confusion in size calling of the microsatellite bands.

The microsatellites that were inconsistent in their inheritance pattern, were difficult to amplify or showed poor levels of polymorphism were dropped. Finally 10 loci were selected from a total of 20 and were further analysed on 72 other individuals (these individual were used in a breeding programme, which is discussed later in section 3.2.2 of this chapter). Table 3.2 lists these microsatellites along with their primer sequence, annealing temperature and Genbank Accession number. The genotype data were analysed using the programme CERVUS. Each of the ten selected loci was characterized for the following parameters:

- Level of polymorphism in terms of number of alleles
- Allele frequency distribution
- Level of informativeness in terms of effective number of alleles, heterozygosity and polymorphic information content (PIC)
- Exclusion power (average probability at a locus of excluding a parent from parentage of a given offspring)
- Hardy-Weinberg equilibrium
- Null allele probability

Two sets of pentaplex PCRs were optimized for simultaneous amplification of the markers. Multiplex PCR was optimized according to the method described in Chapter 2 in section 2.5.

Table 3.2 List of silver carp microsatellite markers taken for characterization

Primer		Sequence	Ta*	Repeat motif	Accession No.
*Hmo*11	*F:*	CTG CTT GAT CAC AGG GTT TG	60	$(GACA)_2(GATA)_8$	AM086451
	R:	CCT TAC AGA TAG ACA GAT ATT CAG			
*Hmo*13	*F:*	AAA CCT GGA AGA TGT TCA CTG AAT	60	$(GACA)_6$	AM086452
	R:	GCG CGA GTG TTT GAA GTC TG			
*Hmo*25	*F:*	TGT GCT GCA TTT TCA CTT CA	60	$(GT)_{13}$	AM086454
	R:	TTC TTA CTA TCC ACA TTT GTT GTA TG			
*Hmo*26	*F:*	GAT TTC AGG CAC ATT GCT TAT CT	60	$(GT)_{40}$ imperfect	AM086455
	R:	GAG CGT TTC TCA TTT GTA CTT ATT TT			
*Hmo*33	*F:*	GTG CAG CAG TAT GTG AAT CAG GAC AC	59	$(GT)_{12}$	AM086458
	R:	GTG CTT CGG GAT ACC ACA CTC TTG			
*Hmo*34	*F:*	GTT CCC TGA GGC TTT ACA A	59	$(GT)_{19}$	AM086459
	R:	GGG TCA TTA TCC TCT CAC TTT			
*Hmo*36	*F:*	ATC GGA GGA GTG CTG TTC AGT CTG GA	63	$(GT)_9$	AM086460
	R:	ACG ATT GTT GCC GAA CGG GTT GAT			
*Hmo*37	*F:*	CAC AGC GGA GGG GCA AAG GTC	63	$(GTGTGA)_3(GT)_3$ $(GC)_2(GT)_{20}$	AM086461
	R:	GGA CGC CGT GTG ACT GGA GAT TTT			
*Hmo*39	*F:*	ACA GTT ATG AGC TAG CAG CAG TTT CT	59	$(GT)_{12}GA(GT)_5$	AM086462
	R:	TAC GTC GTA ATA CCA GTG TAA TAC CC			
*Hmo*40	*F:*	CAG GCA GGC ATC CAC ATA GAG AAT C	63	$(GT)_5(N)_6(GA)_7$ $AA(GA)_4$	AM086463
	R:	AGA AGA AAT CTG ATC GTC ACC TAT GA			

* All the microsatellites were amplified with 1.5 mM $MgCl_2$ concentration

3.2.2 Utility of silver carp microsatellites in parentage assignment

The utility of the silver carp microsatellites in parentage assignment was evaluated first by assessing the resolving power of the loci using the simulation module of the programme CERVUS and then by applying them for actual parentage analysis. For simulation testing the parameters were set as follows:

Cycle (number of offspring to be simulated): 20,000

Candidate parents: 25 and 50 (one sex)

Proportion sampled: 1.000 (default)

Proportion of loci sampled: 1.000 (default)

Rate of typing error: 0.010 (default)

The simulation results were obtained at two confidence limits: strict (95%) and relaxed (80%). Based on these parameters, CERVUS predicts the rate of assignment of parents

(of one sex) under two situations: when the parents of other sex are known and when the other parents are unknown. CERVUS analyses the resolution power of the microsatellites based on likelihood approach.

The ten microsatellites were applied for actual parentage assignment on 3 sets of samples, each consisting of potentially 48 full-sib families, 12 paternal half-sib families and 12 maternal half-sib families. The potential parents for all three sample sets were known. These families of offspring were created in a partial factorial mating (Figure 3.1) between 12 sires and 12 dams in each of the 3 sets (total 72 parents in three sets).

		DAMS											
		D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12
	S01	■	■				■				■		
	S02		■	■				■				■	
	S03			■	■				■				■
S	S04	■			■	■				■			
I	S05		■			■	■				■		
R	S06			■			■	■				■	
E	S07				■			■	■				■
S	S08	■				■			■	■			
	S09		■				■			■	■		
	S10			■				■			■	■	
	S11				■				■			■	■
	S12	■				■				■			■

Figure 3.1: Design of partly factorial mating used for producing full- and half-sib families (shaded cells indicate matings between these sires and dams).

These families were later used for estimating heritability of growth traits in silver carp, which will be discussed in Chapter 4. Since the breeding design is more relevant to the heritability studies it will be discussed in greater detail in Chapter 4. In this chapter, only the success of parentage assignment will be discussed. The exclusion method by programme FAP was used for actual parentage assignment. Before the assignment, the FAP prediction module was used to check the resolving power of the 10 microsatellites for the specific set of families on which the markers would be applied.

3.2.3 Assessing usefulness of the microsatellites in sibship analysis

The usefulness of the silver carp microsatellites in sibship analysis was assessed using the programme KINGROUP (v1.050513). In the absence of parentage information, KINGROUP can reconstruct families if there are some prior ideas about the relationship categories in the sample or population in question. KINGROUP can perform two functions in this regard: (i) it can estimate the likelihood of a pair for having a particular relationship against some other probable relationship(s) using pairwise likelihood ratio and (ii) it can partition individuals in a sample into relationship categories based on the overall likelihood of the pairs belonging to a group.

3.2.3.1 Pairwise likelihood ratio approach

In this method two hypotheses are specified regarding the competing relationship categories: primary and null hypothesis. Every relationship category is associated with two variables, r_p and r_m. These variables define the probabilities that individuals in the pair share an allele by direct descent from their father and mother, respectively. For example, if the presumed relationship is diploid full-sib both the *r* values would be 0.5. If the relationship is paternal half-sibs then r_p is 0.5 and r_m is 0. Given the two hypotheses the program KINGROUP estimates the likelihood of both the hypothesized relationships for a pair of individuals using the *r* values, the population allele frequencies, and the genotype of the two individuals in a dyad and reports the ratio between them (primary/null). The ratio is generally expressed as a log likelihood. The programme can test the significance of the ratio based on simulation. The programme can flag every ratio with "ns" when not significant, "*" for $p<0.05$, "**" for $p<0.01$ and "***" for $p<0.001$ or can provide the *p* value associated with each ratio. If the ratio is found to be significant the primary hypothesis is favoured over the null. In this way the relationship between every possible dyad in a population or sample can be determined.

It is possible to specify more than one relationship (i.e. a range of *r* values) in a hypothesis; then it is called a complex hypothesis. Both primary and null hypotheses can be set as complex. When a complex hypothesis is entered, KINGROUP will find the likelihood for all *r* values within the range and use the highest one in calculating the final ratio. When the one or both of the two hypotheses is set to be complex, the programme cannot perform significance testing. In this case, positive ratios are considered to favour the primary hypothesis and negative ratios favour the null hypothesis.

In the present experiment, the ten microsatellites were used to assess pairwise likelihood ratios for sibship analysis in the following cases:

Case 1: A group of 50 individuals consisting of only full-sibs and unrelated individuals. Five full-sib families were present in the group. Primary hypothesis "full-sibling" was tested against the null hypothesis "unrelated". Two types of errors or misclassification were recorded: (a) when unrelated individuals were misclassified as full-sibs (i.e. false positive) and (b) when full-sibs are misclassified as unrelated (i.e. false negative).

Case 2: A group of 50 individuals consisting only of half-sibs and unrelated individuals. The half-sib individuals came from ten families. These families were produced by the partial factorial mating as discussed in section 3.2.2 and their relationship categories were already known through parentage assignment. The primary hypothesis "half-sibling" was tested against the null hypothesis "unrelated". Misclassification was counted in the same way as discussed above.

Case 3: A group of 77 individuals consisting of full-sibs, half-sibs and unrelated individuals. The primary hypothesis was set to be complex by specifying both "full-

sibling" and "half-sibling" and was tested against the null hypothesis "unrelated". The number of misclassifications was counted as described in *Case 1*.

Case 4: In all the above cases the actual relationships of each pair were known. In *Case 4* the pairwise likelihood approach was applied to determine the pairwise relationship among 72 breeders (the same individuals that were bred in 3 sets using partly factorial mating described in section 3.2.2). In this case the actual relationship between individuals of a pair was unknown, but possible relationship categories were thought to be either full-sib or unrelated (from the available information on the management protocol applied on the broodstock from where these breeders originated). Therefore, the primary hypothesis, "full-sib", was tested against the null hypothesis "unrelated". It would have been ideal if the sib analysis could be performed before the individuals were used in the breeding programme, as in that case matings between full-sibs could have been avoided. Since the breeding was carried out much earlier, this analysis was performed to gain insight into the efficiency of the management practice that is being applied on the stock from where the breeders originated.

3.2.3.2 Partitioning of individuals in relationship categories

The ability of the ten markers to partition individuals into groups of relationship categories was assessed by applying them to ten groups, each consisting of mixtures of full-sibs, half-sibs and unrelated individuals. The primary hypothesis, full-sib, was tested against a complex null hypothesis, half-sib and unrelated. A "descending ratio" algorithm was chosen for the partitioning process. The principle of this algorithm is that it first assigns individuals from pairs with the highest likelihood ratios into groups. Once the first pair is added, then the next highest pair is added and so on until the entire dataset is partitioned.

3.2.4 Assessing usefulness of silver carp microsatellites in relatedness analysis

The objective of this study was to evaluate the efficiency of the ten silver carp microsatellites in relatedness estimation by applying it to individuals of known relationships such as full-sibs, half-sibs and unrelated. For this study allele frequency distributions at 10 loci from 84 individuals were used to simulate 2000 pairs of unrelated individuals, 2000 pairs of half-sibs and 2000 pairs of full-sibs by the programme KINSHIP (ver1.2). The same programme was used to calculate the coefficients of pairwise relatedness of these simulated individuals in each group using the Queller and Goodnight's relatedness estimator (r_{xy}). The estimated relatedness values from each of the three categories were used to create the expected distribution of r_{xy} for the corresponding category. The probability that a pair of individuals of one type will be misclassified as belonging to another category was determined by the method described by Blouin *et al.* (1996). They used the midpoint between the means of two adjacent distributions (e.g. full-sib and half-sib or half-sib and unrelated) as the cut-off value for classification. Therefore, if the distinction was to be made between full-sibs and half-sibs, the cut-off value was 0.375 (midpoint between 0.5 and 0.25, which are the *r* values for full-sibs and half-sibs respectively). Any value that fell to the right of the cut-off value of 0.375 was classified as a full-sib and values to the left as half-sibs. The percentage of the simulated half-sibs that fell to the right of the cut-off value was the expected type I error rate, and the percentage of the full-sibs that fell to the left of the cut off value was the type II error rate. For unrelated individuals and half-sibs the cut-off value was 0.125.

The expected distribution was created with the assumption that there is random mating in the population. The objective was to compare the expected distribution of r_{xy} with

that from actual field data and to determine how accurate these distributions of *r* would be for identifying related individuals. Actual distributions of r_{xy} were produced by analysing 1000 full-sib pairs and 1000 half-sib pairs taken randomly from those created by the partial factorial mating discussed in section 3.2.2. The distribution of unrelated individuals was produced by analysing pairwise r_{xy} for 84 individuals (from which the allele frequency was created) and randomly taking 1000 values.

3.2.5 Cross-species amplification

Silver carp belongs to the family Cyprinidae. All the 10 microsatellite loci were tested for cross amplification in 10 other cyprinid species (Table 3.3). Amplification was tried at different temperatures in the range of 10^{o}C below the annealing temperature used for silver carp and at various $MgCl_2$ concentrations (in the range of 1.5 mM to 3.5 mM). The gradient PCR was used to choose the best temperature and $MgCl_2$ concentration.

Table 3.3: List of species used in cross-species amplification of silver carp microsatllite markers

Scientific name	Common name	Native countries
1. *Labeo rohita*	Rohu	Pakistan, India, Bangladesh, Myanmar, Nepal
2. *Catla catla*	Catla	Pakistan, India, Bangladesh, Myanmar, Nepal
3. *Cirrhinus cirrhosus*	Mrigal	Pakistan, India, Bangladesh, Myanmar, Nepal
4. *Labeo fimbriatus*	---	Pakistan, India, Bangladesh, Myanmar, Nepal
5. *Aristichthys nobilis*	Bighead carp	China
6. *Mylopharyngodon piceus*	Black carp	Amur river basin to southern China
7. *Ctenopharyngodon idella*	Grass carp	China and eastern Siberia
8. *Barbados gonionotus*	Silver barb	Thailand, Indonesia
9. *Cyprinus carpio*	Common carp	---
10. *Tor khudree*	Deccan mahseer	India and Sri Lanka

3.3 Results

3.3.1 Characterization of silver carp microsatellites

3.3.1.1 Polymorphism, informativeness and exclusion power

A summary of the results of CERVUS analysis of ten microsatellite markers is shown in Table 3.4. The number of alleles per locus ranged from 5-16 with a mean of 8.5. Allele frequencies were highly variable at all loci (Figure 3.2). Effective number of alleles (A_e), expected heterozygosity (H_e) and polymorphic information content (PIC) are considered the indicators of informativeness of a marker. All of these variables are heavily dependent on the allele frequency distribution. A marker with a larger number of alleles may have lower A_e, H_e and PIC values than other loci due to a skewed distribution of allele frequency. This phenomenon was also observed in the silver carp markers analysed in the present study. For instance, the locus *Hmo*13 showed higher values for A_e, H_e and PIC compared to *Hmo*11 although the former had a lower number of alleles. *Hmo*36 showed much lower A_e, H_e and PIC values compared to *Hmo*13 and *Hmo*25 even though they had the same number of alleles. Generally a locus having $H_e \leq 0.5$ is not considered very informative. Among the markers analysed in the present study, all except *Hmo*36 had H_e value greater than 0.5. The exclusion powers of a given locus have been shown to be more correlated with the informativeness parameters compared to the actual number of alleles (Table 3.5 and Figures 3.3a to 3.3d).

Table 3.4: Summary of analysis of selected silver carp microsatellite loci

Locus	A	N	Freq	Allele Size range	A_e	H_o	H_e	PIC	Excl(1)	Excl(2)	HW	Null freq
*Hmo*11	7	84	0.52	142-166	2.90	0.607	0.662	0.619	0.256	0.433	NS	0.0439
*Hmo*13	6	84	0.38	136-168	3.64	0.512	0.729	0.679	0.313	0.487	NS	0.1628
*Hmo*25	6	82	0.46	135-146	3.06	0.720	0.677	0.619	0.253	0.419	NS	-0.0356
*Hmo*26	13	84	0.27	145-238	5.07	0.869	0.807	0.775	0.441	0.617	NS	-0.0408
*Hmo*33	11	82	0.33	86-126	4.34	0.829	0.775	0.738	0.391	0.569	NS	-0.0408
*Hmo*34	8	84	0.59	114-132	2.67	0.655	0.629	0.606	0.248	0.439	NS	-0.0244
*Hmo*36	6	84	0.68	208-221	1.95	0.440	0.491	0.447	0.127	0.275	NS	0.0837
*Hmo*37	16	84	0.20	148-192	7.63	0.905	0.874	0.856	0.588	0.741	NS	-0.0229
*Hmo*39	7	84	0.39	118-143	4.11	0.774	0.761	0.725	0.372	0.553	NS	-0.0159
*Hmo*40	5	84	0.56	208-240	2.35	0.583	0.578	0.503	0.174	0.310	NS	-0.0193
Mean	**8.5**				**3.78**	**0.690**	**0.700**	**0.660**				

A= number of alleles
N= number of individuals typed
A_e= effective number of alleles
Freq= frequency of commonest allele
H_o = observed heterozygosity
H_e = expected heterozygosity
PIC = polymorphic information content
Excl (1) = average exclusion probability when both parents are unknown
Excl (2) = average exclusion probability when 1 parent is unknown
HW = Hardy-Weinberg equilibrium test
NS= not significant
Null frequency: see section 2.8.5 for explanation

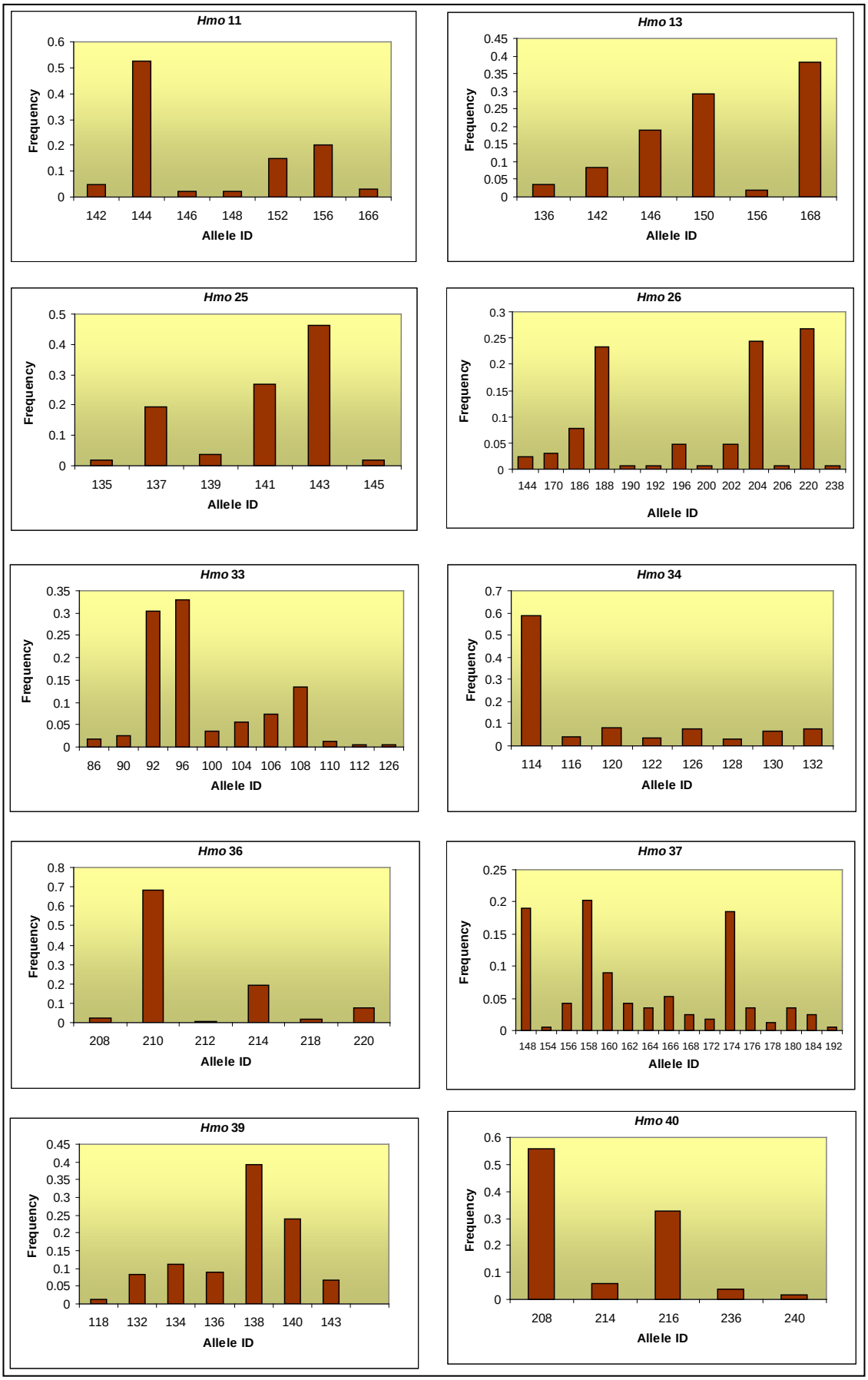

Figure 3.2: Allele frequency distribution at different silver carp microsatellite loci

Table 3.5: Correlation matrix of different parameters of silver carp microsatellite loci

	A	Ae	He	Ho	PIC	Excl(1)	Excl(2)
A	1						
Ae	0.921	1					
He	0.804	0.908	1				
Ho	0.838	0.81	0.855	1			
PIC	0.842	0.919	0.994	0.859	1		
Excl(1)	0.913	0.98	0.966	0.86	0.977	1	
Excl(2)	0.897	0.952	0.975	0.863	0.991	0.993	1

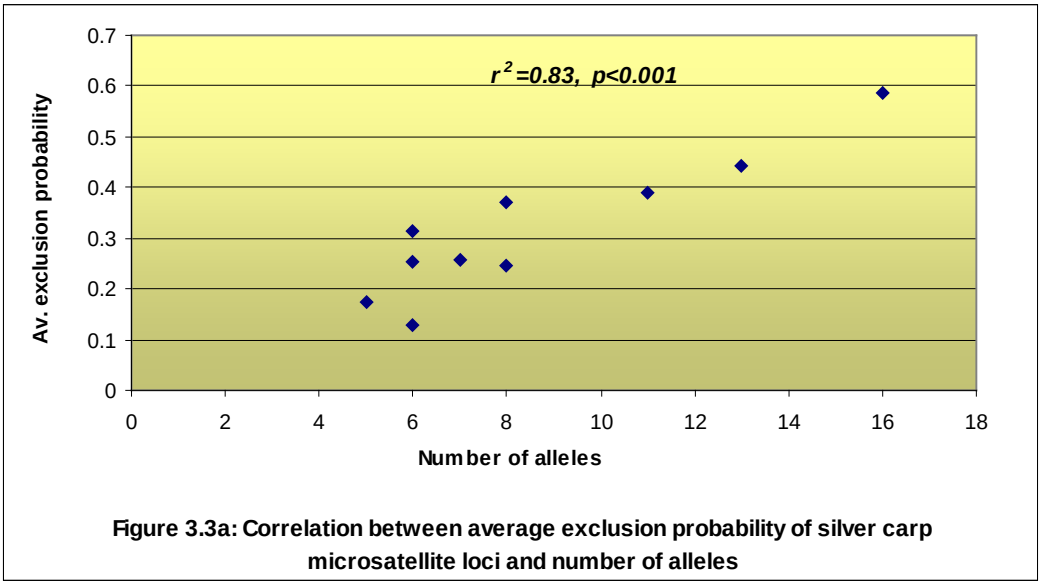

Figure 3.3a: Correlation between average exclusion probability of silver carp microsatellite loci and number of alleles

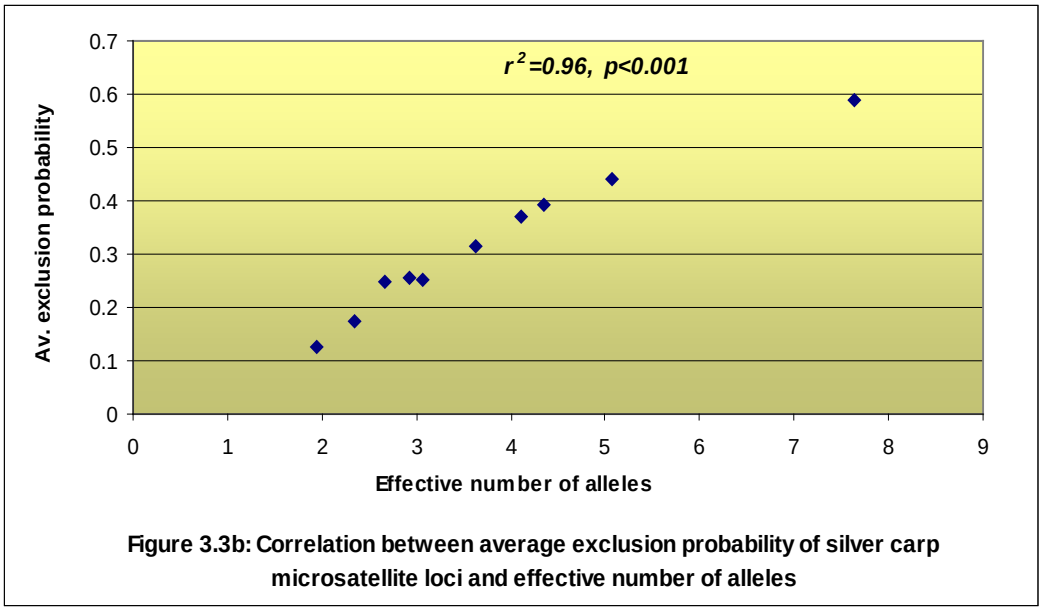

Figure 3.3b: Correlation between average exclusion probability of silver carp microsatellite loci and effective number of alleles

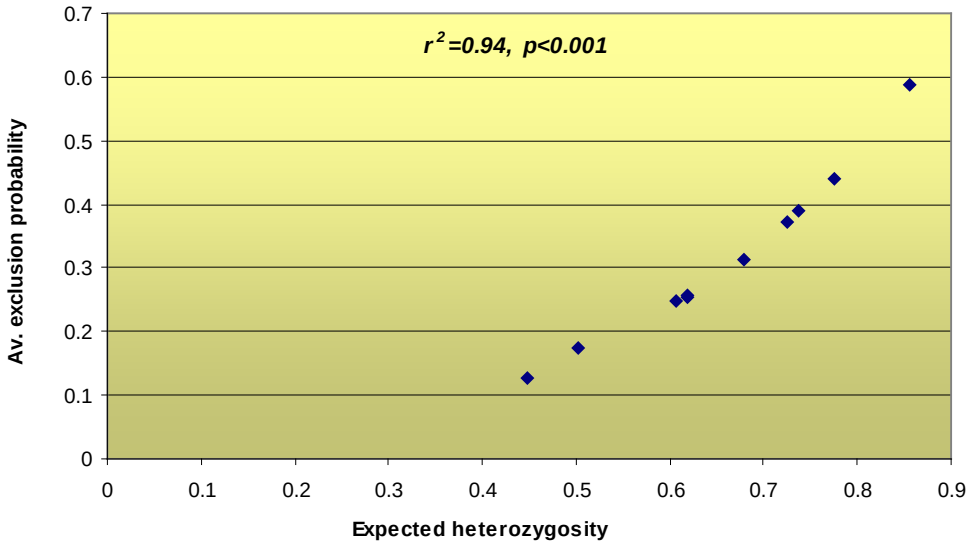

Figure 3.3c: Correlation between average exclusion probability of silver carp microsatellite loci and expected heterozygosity

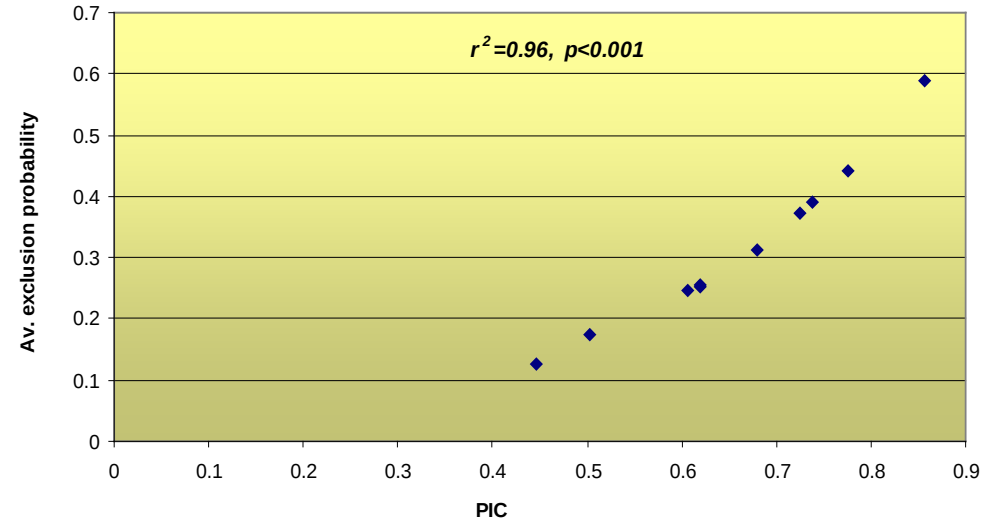

Figure 3.3d: Correlation between average exclusion probability of silver carp microsatellite loci and their PIC values

3.3.1.2 Hardy-Weinberg Equilibrium and null allele prediction

All the microsatellite loci analysed showed Hardy-Weinberg equilibrium (HWE) (Table 3.4) indicating the absence of major sub-grouping in the population, major linkage and null allele problems. Estimated null allele frequency has also been reported in this table. Three loci, namely *Hmo*11, *Hmo*13 and *Hmo*36 showed positive null allele frequencies even though these did not cause deviations from HWE in any of these loci. Particularly, *Hmo*13 showed a rather large null prediction and hence was used with caution during the parentage assignment. The remaining seven microsatellite loci showed slightly negative null frequencies, indicating the presence of excess heterozygotes.

3.3.2 Optimization of multiplex PCR

Two pentaplex PCRs were optimized during the present study. Tables 3.6a and 3.6b show the primer combinations of the two multiplex PCRs and Table 3.6c presents optimized PCR reaction and thermocycling condition. Variable concentrations of the primers had to be used to get similar intensity of the amplified products across the loci during fragment analysis. Genotype images of the two multiplexes are shown in Figures 3.4a and 3.4b. The major features of multiplex PCRs were the use of greater concentrations of all the PCR reagents compared to singleplex reaction. In particular, the Taq DNA polymerase concentration was elevated to three times (0.15 U per reaction) the level used in singleplex PCR (0.5 U per reaction). This helped to reduce the amplification of spurious bands. PCR buffer II (ABgene) which contained KCl was used instead of the more communally used buffer IV (the compositions of buffers II and IV are in sections 2.5 and 2.4 of Chapter 2). KCl is believed to reduce differential amplification across different loci by facilitating primer annealing (Fishback *et al.* 1999; Wallin *et al.* 2002). The multiplex worked well under the same thermocycling conditions as used for singleplexes and hence this was not changed.

Table 3.6a: Primer information on the set A multiplex PCR

Locus	Size range (bp)	Fluorescence	Primer concentration (pmol/µl)*	Individual T_a (^oC)	T_a used (^oC)
*Hmo*25	135-146	FAM	0.06	60	
*Hmo*26	145-238	HEX	0.25	60	
*Hmo*36	208-221	FAM	0.07	63	60
*Hmo*37	148-192	TET	0.09	65	
*Hmo*39	119-143	TET	0.09	59	

***** denotes primer concentration for both forward and reverse primers

Table 3.6b: Primer information on the set B multiplex PCR

Locus	Size range (bp)	Fluorescence	Primer concentration (pmol/µl)*	Individual T_a (^oC)	T_a used (^oC)
*Hmo*11	142-166	FAM	0.05	60	
*Hmo*13	136-168	TET	0.045	60	
*Hmo*33	86-126	FAM	0.05	59	60
*Hmo*34	114-132	HEX	0.20	59	
*Hmo*40	208-240	TET	0.06	63	

***** denotes primer concentration for both forward and reverse primers

Table 3.6c: PCR reaction and thermocycling conditions for multiplex PCR

PCR reaction condition (amount per PCR reaction)		**Thermocycling condition**		
Buffer:	x 2	95^{o}C	2 min 30 sec	
dNTP:	280µM	94^{o}C	50 sec	30 cycles
$MgCl_2$:	2mM	60^{o}C	50 sec	
Taq:	1.5U	72^{o}C	1 min	
DNA:	100ng	72^{o}C	35 min	
		10^{o}C	hold	

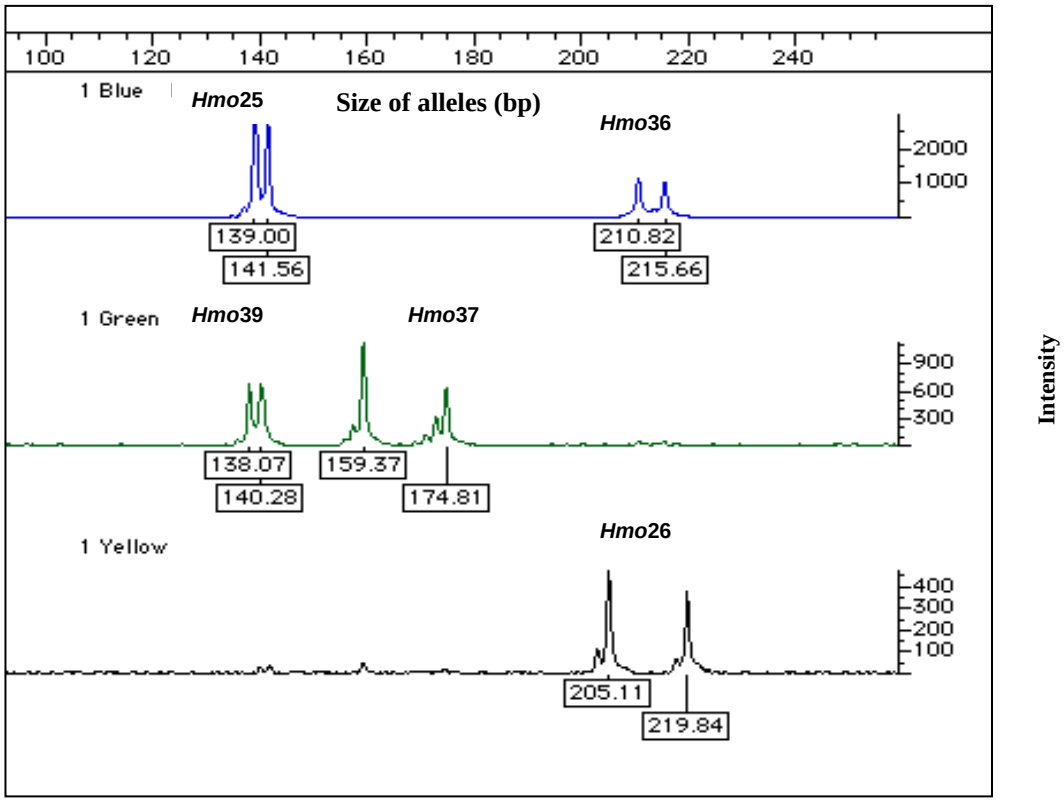

Figure 3.4a: Genotype image of amplifications at different loci by multiplex PCR set A

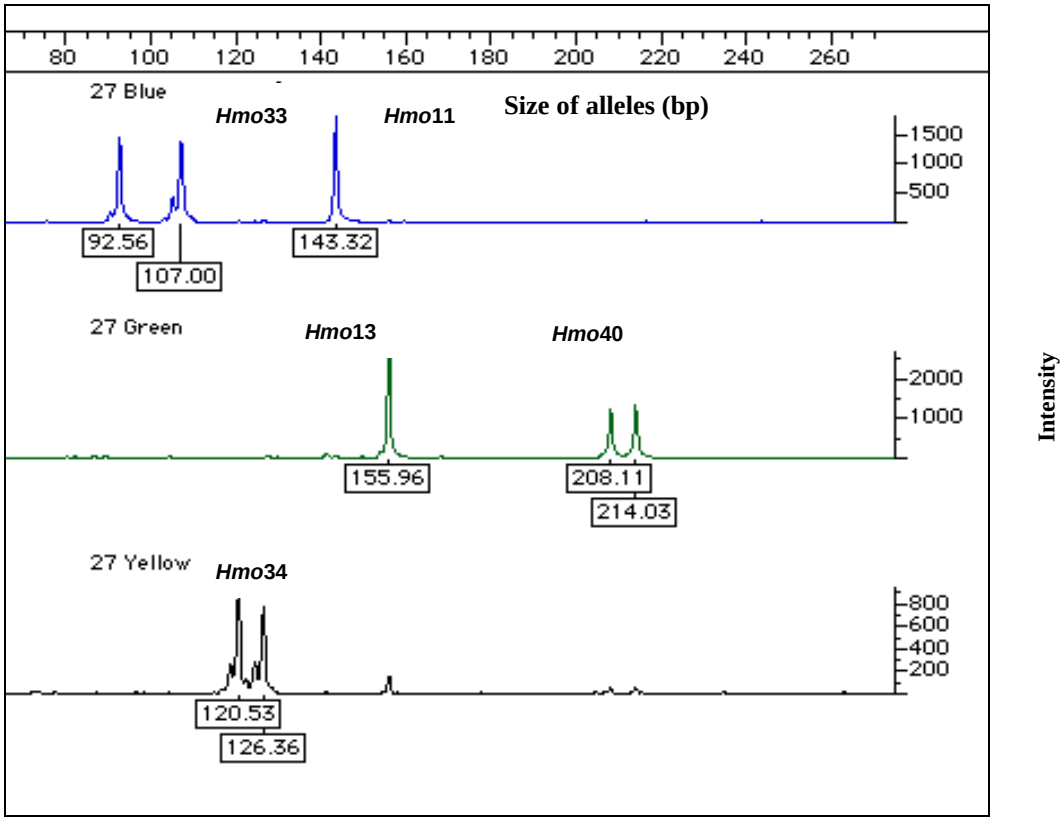

Figure 3.4b: Genotype image of amplifications at different loci by multiplex PCR set B

3.3.3 Silver carp microsatellites in parentage analysis

3.3.3.1 Resolving power of the microsatellites

The resolving power of the ten microsatellites was first tested by CERVUS likelihood based simulation involving 25 parents or 50 parents of a single sex and the result is presented in Table 3.7. If one parent for each offspring is already known, then the ten microsatellites can achieve 100% assignment for both 25 and 50 parents. If, however, neither parent is known the resolving power of the available markers drops drastically, giving 75% success for 25 parents and only 58% for 50 parents. The resolving power of the markers was also determined by the FAP exclusion based process for specific structured breeding on which the microsatellites were ultimately applied for actual assignment (Table 3.8). In this case, the ten markers gave a very high rate of assignment prediction (average 98.71%) for the structured breeding with 12 sires and 12 dams.

Table 3.7: Resolving power of silver carp microsatellites by likelihood based simulation

Number of candidate parents	**When 1 parent unknown**		**Both parents unknown**	
	80% confidence	95% confidence	80% confidence	95% confidence
25	100%	100%	100%	75%
50	100%	100%	99%	58%

3.3.3.2 Actual parentage assignment

The actual parentage assignment was performed on a total 331 offspring, obtained from three sets of partial factorial mating where each set had 12 sires and 12 dams (see Figure 3.1). 319 offspring (96.37%) could be assigned to a single parental pair whereas the average predicted assignment success (for three sets) was 98.71%. The actual assignment rate varied in the three sets with 100% success rate in Set A, 94.87% in Set B and 94% in Set C. From the 12 unmatched offspring, 8 (2.41%) matched to two

families and 4 (1.2%) matched with more than two families (Table 3.8). Since locus *Hmo*13 showed a high prediction of null alleles, parentage assignment was first carried out without this locus and then it was applied only on unassigned offspring. However, for Set B parentage assignment without *Hmo*13 was very low (only about 48%) and hence the marker was applied on all the individuals in this set. During FAP analysis, an allele mismatch tolerance was set at 4. Thus manual correction was done afterwards wherever necessary.

Table 3.8: FAP based parentage prediction and actual parentage assignment result

	Set A	**Set B**	**Set C**	**Total**
Prediction for assignment success	99.58%	97.82%	98.73%	98.71%
Total number individuals in set	114	117	100	331
Assigned to a single family (after using *Hmo*13)	114 (100%)	111 (94.87%)	94 (94%)	319 (96.37%)
Assigned to 2 families	0	3 (2.56%)	5 (5%)	8 (2.41%)
Assigned to more than 2 families	0	3 (2.56%)	1 (1%)	4 (1.2%)

3.3.3.3 Null alleles

During parentage assignment with the programme FAP, a certain level of mismatches was tolerated between parents and offspring (max. 4 alleles out of 20 alleles from 10 loci). A mismatch at a locus was considered to be due to a null allele when one of the parents of an offspring was homozygous and the offspring was also homozygous but for a different allele which was present in the other parent (Dakin and Avise 2004). For instance, in an $A_1A_1 \times A_1A_2$ cross, a null allele in the individual A_1A_1 can be detected when an A_2A_2 genotype is observed in the progeny. This putative homozygous has

actually a $0A_2$ genotype and the homozygous parent A_1A_1 would then be a heterozygous $0A_1$ (the example has been drawn from Castro *et al.* 2004).

In the current study, the presence of null alleles was confirmed in *Hmo*13. In Set B it affected 6 parents (25% of the parents in that set) and 14 offspring from these parents. Two other loci, *Hmo*26 and *Hmo*39 also showed null alleles (reported in Table 3.9) even though CERVUS prediction did not produce any positive null allele frequency for these two loci.

Table 3.9: Probable null alleles in the silver carp microsatellites

Locus showing null allele	Number of parents affected	Number of offspring affected
*Hmo*13	6	14
*Hmo*26	1	6
*Hmo*39	4	7

3.3.4 Silver carp microsatellites in sibship analysis

3.3.4.1 Using pairwise likelihood ratio

The result of the pairwise likelihood ratio approach in determining the relationship of a dyad from a few predefined categories is summarized in Table 3.10. The markers were found to be reasonably effective when the distinction was to be made between full-sib and unrelated (*Case 1*). Only 4% of the full-sibs were misclassified as unrelated. However, the rate of false positives i.e. misclassification of unrelated as full-sib was quite high (11.1%). This indicated that the parental population was probably inbred to a certain extent.

The markers showed much lower accuracy when half-sibs were to be distinguished from unrelated (*Case 2*). In 42.15% of cases a false negative result was observed, misclassifying half-sibs as unrelated. In contrast, the misclassification of

Table 3.10: Result of sibship analysis by pairwise likelihood method using 10 silver carp microsatellites

Sample structure	Number of individuals	No. of pairs	Primary (H_p) and null hypotheses (H_o)	Misclassification/ errors in reconstruction
Case 1: Full-sibs + unrelated	50	Total pairs: 1225 Unrelated: 1000 Full-sib: 225	H_P: Full-sibling H_o: Unrelated	*Unrelated as full-sibs:* 111 out of 1000 pairs (11.1%) *Full-sib as unrelated*: 9 out 225 pairs (4%)
Case 2: Half-sibs + unrelated	50	Total pairs: 1225 Unrelated: 1123 Half-sib: 102	H_P: Half-sibling H_o: Unrelated	*Unrelated as half-sibs:* 57 out of 1123 pairs (5.07%) *Half-sib as unrelated:* 43 out 102 pairs (42.15%)
Case 3: Full-sibs + half-sibs + unrelated	77	Total pairs: 2926 Unrelated: 2293 Full & half-sib: 633 ➢ Full-sib: 146 ➢ Half-sib: 487	H_P: Full-sibling H_P: Half-sibling H_o: Unrelated	*Unrelated as sibling:* 302 out of 2293 pairs (13.2%) *Siblings as unrelated:* 116 out 633 pairs (18.33%)

unrelated as half-sibs was only 5.07%, much lower compared to that observed in the above case involving full-sib groups. Since hatchery populations are expected to involve different relationship categories the markers were applied on a more realistic situation involving full-sibs, half-sibs and unrelated individuals (*Case 3*). Rather than distinguishing between and among every possible relationship, in this case the objective was to tell if the individuals were unrelated or sibs. This would allow matings between siblings to be avoided. The rate of false positives (unrelated as siblings) was observed to be 13.2% and rate of false negatives (siblings as unrelated) was 18.33%. Even though this population consisted of a much greater proportion of half-sib (487 pairs) compared to full-sibs (146 pairs), the rate of false positives was observed to be much lower than was expected from the observation in *Case 2*. The probable reason for having this result is that in *Case 3* the null hypothesis was complex. When one of the hypotheses is complex, the programme KINGROUP does not calculate the significance of the likelihood ratio. Therefore, any positive result was considered to favour the primary hypothesis even if it was not significant. On the contrary in *Case 1* and *Case 2* many of the positive results might have been non-significant and thus have been rejected in favour of the null hypothesis.

After assessment of the silver carp microsatellite markers for sib-ship analysis by applying to pairs with known relationship, the markers were applied to a group of 72 silver carp breeders (these breeders were used in the mating discussed in section 3.2.2) where the relations between each pairs were unknown (*Case 4*). These breeders presumably originated from a well managed broodstock. In that population breeding was practiced by single pair mating, for broodstock replacement purpose. Thus the likely relationship between pairs would be full-sib and unrelated. When analysed for full-sibship among the breeders (against the null hypothesis, unrelated), 8.61% of the

pairs (220 pairs out of total possible pair combination of 2556) showed a significant result indicating a possible full-sib pair. However, from previous case studies it should be recognised that there would be certain percentage of error in classification in both directions (i.e. full-sib as unrelated and unrelated as full-sib). Table 3.11 shows the result of sib analysis (without adjustments for misclassification) at different significance levels. The actual breeding record showed that in 16 cases mating between these full-sibs took place as a matter of chance (Table 3.12). From these full-sib matings a total of 22 offspring were produced (this information could be gathered as microsatellites were applied for parentage analysis). The survival and growth performance of the progeny of full-sib parents were compared with those of progeny group of unrelated parents, to find out if there was any interpretable difference in the performance (shown in Table 3.12). The table shows that the mean weight and length of the progeny of full-sib parents was higher than those of the progeny of unrelated parents. However, the mean number of offspring produced by full-sib parents (1.38) was lower than that of unrelated parent pairs (2.38).

Table 3.11: Summary of the analysis of sib-ship of silver carp breeders*

Total number of possible pair combination (from 72 individuals)	2556
Putative full-sib pairs	220 (8.61%)
Classified according to significance level	
➢ (≥1 – <5%)	130
➢ (≥0.1 – <1.0%)	58
➢ (< 0.1%)	32

* Without adjustments for misclassification

Table 3.12: Information about the survival and growth performance of progeny from full-sib parents compared to the performance of progeny from unrelated parents

Actual breeding due to chance (in pairs)	16
Total number of offspring produced by full-sib families	22
No. of offspring produced by full-sib parents	0-6
Mean (± SD) number of offspring per full-sib family	1.38 ± 1.78
Mean (± SD) number of offspring per unrelated family	2.38 ± 2.27
Mean weight (± SD) of the offspring from full-sib parents	432.50 ± 42.05 (g)
Mean weight (± SD) of the offspring from unrelated parents	403.18 ± 81.52 (g)
Mean length (± SD) of the offspring from full-sib parents	34.20 ± 0.95 (cm)
Mean length (± SD) of the offspring from unrelated parents	33.17 ± 2.11 (cm)

3.3.4.2 Using group partitioning approach

Group partitioning was applied in ten groups of fishes, each of which consisted of full-sibs inside half-sib families. Some groups also included a few unrelated individuals. Since, here the objective was to sort individuals into subgroups (i.e. full-sibs) from broader groups (half-sibs), the null hypothesis (complex) was set as half-sib and unrelated and the primary hypothesis as full-sib. The result of group partitioning is summarized in Table 3.13. Errors in partitioning are described under two headings: wrong partition and extra partition. If an error was due to misclassification of a member of a group to another group, it was reported under “wrong partition” and when a group was split into more than one group it was placed under “extra partition”. Table 3.13 shows that except in the first group, in all other cases a certain percentage of errors, ranging from 6% to 45% occurred. The cases of wrong partitioning were generally low. In most cases the error was due to splitting the members of a group into several groups.

Table 3.13: Result of pedigree reconstruction by group partitioning method using silver carp microsatellite markers

Groups	No. of individuals	Hypotheses	Expected number of partitions	Observed number of partitions	Errors in partitioning		
					Wrong partition	Extra partition	Error rate
➢ Group 1	11		6	6	➢ None	➢ None	0%
➢ Group 2	17		6	7	➢ 1 individual	➢ 1 FS group split into 3 groups	23.53%
➢ Group 3	11		3	3	➢ 2 individuals	➢ None	18.18%
➢ Group 4	11	H_P: Full-sib (FS)	3	5	➢ None	➢ 2 extra groups each with 1 individual	18.18%
➢ Group 5	17	H_o: Half-sib (HS)	4	5	➢ 1 individual	➢ 1 individual placed in 1 extra group	11.76%
➢ Group 6	17	H_o: Unrelated	6	7	➢ None	➢ 1 individual placed in 1 extra group	5.89%
➢ Group 7	9		2	3	➢ None	➢ 1 individual in 1 extra group & 1 group split into 2 groups	11.11%
➢ Group 8	12		5	7	➢ None	➢ 1 individual in 1 extra group & 1 group split into 2 groups	33.3%
➢ Group 9	20		2	6	➢ None	➢ 2 groups each partitioned into 3 groups	45%
➢ Group 10	13		4	8	➢ None	➢ 4 FS in 4 extra groups	30.76%
						Mean error rate	19.2%

3.3.5 Silver carp microsatellites in relatedness analysis

The mean (± SD) of r_{xy} distributions for three relatedness classes from simulated and actual data are shown in Table 3.14. The means of the r_{xy} of simulated individuals in 3 classes corresponded to the expected values of 0, 0.25 and 0.5 for unrelated, half-sibs and full-sibs respectively. For both predicted and actual distributions the standard deviation was very high, in the range of 0.17 to 0.21. Figure 3.5 compares the distribution of r_{xy} in three classes between simulated and actual data. For unrelated and full-sib the predicted and actual distributions were very similar although the spreads of the distributions were greater in actual data. Unlike full-sibs and unrelated, in half-sibs the actual distribution has shifted towards the left, giving a lower mean value of r_{xy} (0.207 ± 0.20). This means that more half-sib pairs will be misclassified as unrelated but less will be misclassified as full-sibs in the actual dataset compared to the predicted rates of misclassification. Predicted and actual rates of misclassification between different classes are reported in Table 3.15. As expected, about 10% more half-sib pairs were misclassified as unrelated and about 5% less half-sib pairs were misclassified as full-sib in real data. For other cases the actual misclassification rate was similar or close to the predicted rates.

Table 3.14: Mean and standard deviation of r_{xy} distribution of different relationship categories in simulated and real individuals

	Number of individuals	Distribution of r_{xy}		
		Unrelated	**Half-sibs**	**Full-sibs**
Simulated	2000	0.003 ± 0.186	0.248 ± 0.183	0.495 ± 0.173
Real	1000	0.01 ± 0.211	0.207 ± 0.2	0.449 ± 0.194

Table 3.15: Predicted and actual misclassification rate using r_{xy} in different relatedness classes

True relationship	Misclassified as	Cut-off point	Misclassification rate	
			Predicted	**Actual**
Unrelated	Half-sib	>0.125 & <0.375	22.2%	23.2%
Half-sib	Unrelated	< 0.125	24.85%	35.3%
Half-sib	Full-sib	> 0.375	25.35%	20.7%
Full-sib	Half-sib	<0.375 & >0.125	21.55%	19.3%
Unrelated	Full-sib	>0.375	2.55%	5.1%
Full-sib	Unrelated	<0.125	2.3%	4.3%

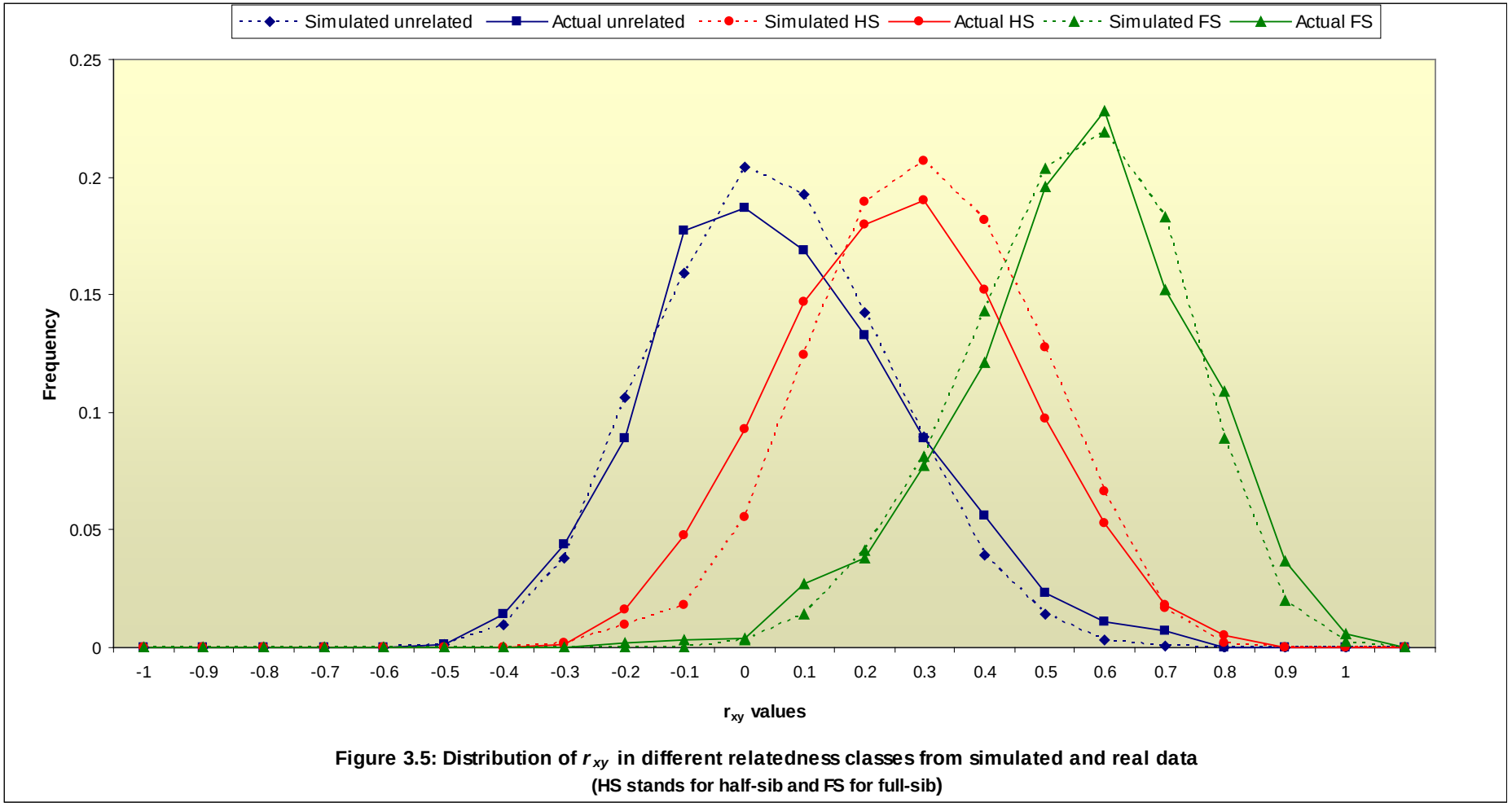

Figure 3.5: Distribution of r_{xy} in different relatedness classes from simulated and real data
(HS stands for half-sib and FS for full-sib)

3.3.6 Cross-species amplification of silver carp microsatellites

Table 3.16 shows the result of cross species amplification of silver carp microsatellite in ten other cyprinid species. The table reports only those cases where the products of cross-amplification were similar in size to the product from silver carp. All markers except *Hmo*39 and *Hmo*40 amplified in every species. In three species, namely bighead carp, grass carp, and black carp, all of the markers amplified. Out of 100 marker-species combinations, in 87 cases successful amplification was recorded. In 63 cases the PCR was successful in the original PCR and thermocycling conditions that were used for the marker in silver carp. In other cases annealing temperature (T_a) and/or $MgCl_2$ concentration had to be adjusted to get a PCR product. In 5 cases more that one band was observed on agarose gel. Due to shortage of time it was not possible to investigate the polymorphism of the markers that amplified in other species.

Table 3.16: Cross amplification of 10 silver carp microsatellites in 10 other cyprinid species

Loci	*L. rohita*	*C. catla*	*C. cirrhosus*	*L. fimbriatus*	*A. nobilis*	*M. piceus*	*C. idella*	*B. gonionotus*	*C. carpio*	*T. khudree*
*Hmo*11	+	+	+	+	+	+	+	+	+	+
*Hmo*13	+	+	+	+	+	+	+	+	+	+
*Hmo*25	60, 2.5	60, 2.5	60, 2.5	60, 2.5	+	+	57, 1.5	60, 2.5	60, 2.5	60, 2.5
*Hmo*26	+	+	+	+	57, 2.5	+	+	60, 2.5	53, 1.5	57, 1.5
*Hmo*33	50, 2.0	50, 2.0	50, 2.0	59, 2.5	+	+	+	+	44, 2.0*	56, 2.5
*Hmo*34	+	+	+	+	+	+	+	+	+	+
*Hmo*36	+	+	+	+	+	+	+	59, 1.5*	+	59, 1.5*
*Hmo*37	+	+	+	+	+	+	+	+	+	+
*Hmo*39	—	46, 2.0	—	—	49, 2.0*	50, 2.5	50, 2.5*	—	—	—
*Hmo*40	—	—	—	—	+	+	+	—	—	—

— No amplification
\+ Amplification in original conditions for silver carp; for other cases T_a & $MgCl_2$ concentration mentioned
* More than one band observed on agarose gel

3.4 Discussion

3.4.1 Characterization of silver carp microsatellites

Ten silver carp microsatellite markers were selected from a panel of 20 markers based on their conformity to Mendelian transmission, polymorphism and reproducibility and were characterized for applications in aquaculture genetic management. The selected markers showed moderate levels of polymorphism in terms of actual number of alleles (5 to 17 alleles). Allele frequency distribution, however, is the key factor in determining the informativeness of a marker and hence its usefulness in pedigree analysis. The exclusion power of a marker in parentage analysis, reaches its maximum value when all alleles have the same frequency (Villanueva *et al.* 2002). Such a situation, however, is rarely found in practice. The informativeness of most of the markers in the present study has also suffered due to highly uneven allele frequency distribution. One microsatellite showed a heavy presence of null alleles. Dakin and Avise (2004) showed that null alleles generally have little impact on the average exclusion probabilities of a set of markers but in actual parentage assignment they can introduce substantial errors. Therefore, it is preferable to exclude any microsatellites showing severe null allele problems or else their primers can be redesigned for the focal population.

Two multiplex PCR reactions were optimized in this study for rapid genotyping of a large number of samples. Multiplex PCRs not only accelerated the genotyping process but in certain cases produced clearer genotypes by reducing spurious bands and stutters compared to that obtained in singleplex PCR. Deciding which set of loci are to be coamplified and standardizing the relative concentrations of the primers were the major steps towards the optimization of the multiplex systems. The general competence of the markers was assessed by analysing the free energy level (ΔG) between each pairs of primers (also discussed in section 2.5). The greater the ΔG values are, the less

competent the primers and vice versa. However, analysis of the ΔG alone may not ensure successful amplification of all loci. It may still require a great deal of trial and error to optimize a multiplex system. In the present study the ΔG approach gave a fairly good assessment about which microsatellites should be amplified together. Variable concentrations of primers had to be used to get similar signal strength and amplification specificity across the loci. Great difference in intensity between products can seriously hamper the size calling process in two ways. First, weak signals may not be picked up by the automated genotyping system. In this case it will require a great deal of manual correction. Second, too strong signals may lead to offscale data and fluorescence bleedthrough ("pull-up") making size calling difficult. Differential amplification of the participating loci in a multiplex system may be caused by competition or interaction between the primers and by the fluorescent dye attached to the primer. The primers were labelled with either of three phosphoramadite fluorescent dyes: FAM (blue), TET (green) and HEX (yellow). FAM produces the brightest fluorescence whereas HEX produces the weakest. This is largely due to dye excitation efficiency (Wallin *et al.* 2002). The closer the dye's excitation maximum is to the laser's wavelength(s), the brighter the emitted fluorescence. Therefore, if the same markers are used with different dyes, the concentration of the primers might require fresh adjustment.

3.4.2 Utility of silver carp microsatellites in kinship analysis

Microsatellites have long been established as invaluable tools in kinship analysis in farmed and natural populations. In aquaculture, the importance of pedigree information rests in minimizing inbreeding, maintaining effective population size, maintaining genetic diversity and in selective breeding. The present study assessed the utility of a set of newly developed silver carp microsatellites in pedigree analysis as a means of genetic management in this species. Different approaches to pedigree analysis were explored to

match the need of different circumstances associated with hatchery populations. Which approach to apply in real situations will have to be decided upon based on how much pedigree information is already known about a population and also upon the purpose of the management practice. The most straightforward case of pedigree analysis is the parentage assignment when all the potential parents are known. It is however, quite likely to encounter situations in hatcheries where parental information are only partially known or are totally missing because the stock was left unmanaged or the stock has recently been procured from another source. If there are some prior ideas about the type of relationship categories involved in such a stock, then pairwise relationships can be estimated or group partitioning can be performed using the likelihood ratio test. The pairwise likelihood approach can be useful in avoiding mating between relatives. Group partitioning on the contrary, should have broader utility in maintaining sib families for purposes such as estimation of genetic parameters or in selective breeding. One major problem associated with both the pairwise and group-partitioning is that possible relationship categories need to be specified which may not be obvious in many situations. For such situations pairwise relatedness analysis was explored as a means of kinship analysis.

In the present study, an average of about 96% parentage assignment success was attained with 10 loci when applied to a partial factorial mating design involving 12 sires and 12 dams and producing 48 full-sib families. In a similar full factorial breeding programme (10 sires and 10 dams and 100 potential families) with rainbow trout Herbinger *et al.* (1995) attained 91% assignment using only 4 or 5 loci. Vandeputte *et al.* (2004) attained a success rate above 95% in a full factorial mating between 10 dams and 24 sires in common carp using 8 microsatellites. In both these instances the numbers of potential families were much larger than that in the present experiment.

Thus although for the setting of the present breeding structure the group of silver carp microsatellites were highly successful in parentage assignment, when compared to other studies their performance cannot be considered very satisfactory. This is due to the fact that markers in the present study were not highly informative as was shown by their effective allele number, expected heterozygosity and polymorphic information content. Villanueva *et al.* (2002) showed by deterministic prediction that with nine 5-allele loci or six 10-allele loci with equifrequent alleles it is possible to assign parentage to 99% of the fish resulting from either 100 or 400 crosses. Although the average number of alleles for the loci used in the present experiment was 8.5, still the resolution power was nowhere near the prediction by Villanueva *et al.* (2002). This is obviously due to highly skewed allele frequency distribution observed across these loci. A simulation based study showed (result in section 3.3.3.1) that if a larger number of parents were involved than in the present experiment, these microsatellites will not be sufficient to obtain a high percentage of assignment.

When applied to estimate the likelihood of a pair of having one relationship over some other, the markers were found to be very effective in discriminating between full-sibs and unrelated individuals. In 96% of the cases full-sib pairs were correctly recognized. However, a larger percentage of unrelated pairs were misclassified as full-sibs (11.1%, as opposed to 4% for full-sib as unrelated). Possible reasons can be the presence of alleles which were identical by state and not by descent or the parental population showed a certain level of inbreeding. Since the main goal of applying this pairwise approach would be to minimize inbreeding, the rate of false negative results (i.e. misclassification of full-sibs as unrelated) will ultimately determine the success. When the discrimination was to be between half-sib and unrelated, the markers behaved poorly, giving correct classification for about 58% of half-sib pairs. In this case, 95% of

the unrelated pairs were correctly classified, indicating a less inbred parental generation compared to the first case. While the number of available markers and their informativeness definitely play important roles in the success of pedigree reconstruction, a more important determinant is how close the IBD co-efficients are for the competing categories (Blouin 2003). In general 10 loci should give a fair discrimination (90%) between parent and offspring and 15 to 20 loci will be required for discriminating full-sibs from unrelated. Around 50 loci might be required for similar power to discriminate half-sibs from full-sibs or unrelated. The result of the present study conformed to this general rule or showed slightly better performance by the markers. The present study showed that if the likelihood ratios are considered based on their positive or negative values rather than their statistical significance, a greater percentage of family reconstruction may be obtained as was observed in a complex population involving full-sib, half-sib and unrelated individuals. The microsatellites were observed to be reasonably effective when used for group partitioning. Correct partition was attained for an average 80% of the individuals. However, this result should be considered with caution as only a few individuals were involved in each sample where the partitioning was performed.

When pairwise likelihood approach was applied on a group of breeders, 8.61% of the pairs appeared to be putative full-sibs. These silver carp breeders have originated from a broodstock which has been maintained and managed for several generations (3-4 generations) in NFRDMP hatchery, in Bangladesh. Silver carp being an exotic species, the broodstock could not be replaced by wild fish from nature. Every generation, broodstock have been replaced by progeny of the previous generation which was bred in a 1:1 ratio to keep the rate of inbreeding to minimum. Considering the fact that, 3 to 4 generations of broodstock replacement has been practiced without procuring fish from

outside, the 8.61% of full-sib pairs appears to be reasonably low. This indicates that the management practice has been quite efficient. Full-sib mating increases the inbreeding level by 25% (Falconer and MacKay 1996). However, comparison of growth and survival performance of progeny from full-sib parents with the performances of progeny from unrelated parents did not show any alarming decrease due to inbreeding (in fact the mean weight and length of progeny of full sib parents were higher). Of course, it must be acknowledged that the number of offspring was very few (only 22 from 16 families) to draw accurate conclusion about the effect of inbreeding.

Silver carp microsatellites were used to estimate pairwise relatedness co-efficients using Queller and Goodnight's (1989) estimator. This estimator requires population allele frequency. If the population allele frequency is not known but is estimated from a sample, Queller and Goodnight's estimator may produce biased results. For the present context allele frequency was determined from genotypic data from 84 individuals, presumed to be unrelated, and was considered as population allele frequency because all the progeny used in this study came from these individuals. To apply the relatedness approach it is important to create a reference distribution of *r* for different relatedness categories so that distributions of pairwise *r* estimated on real data can be assessed against this. The reference population can also be used to predict the error rates of misclassification of pairs. Blouin *et al.* (1996) first described the method of developing reference *r* distribution using population allele frequency, which was later followed by other workers for work on Atlantic salmon (Norris *et al.* 2000), rainbow trout (McDonald *et al.* 2004) and Senegalese sole (*Solea senegalensis*) (Porta *et al.* 2006). In the present experiment the reference distribution was found to be quite effective in comparison to the actual distribution and for predicting the error rate. The actual distribution for full-sib and unrelated groups were in close conformity to the reference

distribution but the distribution of half-sibs showed a leftward shift with a lower mean value of *r*, leading to misclassification of more half-sibs as unrelated. According to McDonald *et al.* (2004) the conformity of the actual distribution of *r* to the reference distribution is more important than the matches of the mean values. As long as the two distributions are identical, any change in mean value should not affect the accuracy of the unrelated pair selection.

The relatedness distributions estimated in the present study showed large standard deviations (0.173-0.211), which are typical of any relatedness studies (Blouin 2003). Standard deviations for relatedness distributions reported in some other works are: 0.114-0.122 in wild mice (*Mus musculus*) (Blouin *et al.* 1996), 0.148-0.189 in rainbow trout (McDonald *et al.* 2004), 0.218-0.250 in turbot (Borrell *et al.* 2004) and 0.147-0.161 in Senegalese sole (Porta *et al.* 2006). While the presence of alleles identical by state (and not by descent) and variable IBD co-efficients across different loci are important reasons for inherently large standard deviation for *r* estimation, the number of marker loci and their informativeness are also important sources of variance (McDonald *et al* 2004). Thus, although the microsatellites used by Blouin *et al.* (1996) had similar levels of H_e (average 0.73) to the microsatellites used in this study (average H_e=0.70), their standard deviation of the *r* estimation was much lower because they used a higher number (20) of loci.

The present study gives only an initial idea about the applicability of the silver carp microsatellites in pedigree analysis. The samples used in the present study came from a well managed hatchery population where the genetic variability has been maintained and inbreeding level has been carefully tackled. It is therefore quite likely that the result observed in the present study will not be reproduced in all cases, especially in

populations which have suffered loss of genetic diversity and inbreeding over several generations. This was demonstrated by McDonald *et al* (2004) as they observed different levels of success in relatedness analysis in three hatchery strains which had different levels of genetic variability. They created reference distributions of *r* for each strain using the strain specific allele frequency. One of these strains had undergone several generations of selective breeding and thus showed a lower level of genetic variation compared to the other two strains. The predicted distribution of *r* for different relationship classes in this strain showed greater overlapping and hence the predicted rate of error was also higher compared to other two strains. Therefore it can be concluded that to understand more about the applicability of the silver carp markers it is important to assess them under wider situations. It is also important to develop more microsatellites for this species to obtain greater precision in kinship analysis. Recently Tong *et al.* (2002) reported successful cross-amplification of 7 microsatellites in silver carp with common carp primers. A number of other silver carp microsatellite sequences are also currently available in GenBank (although yet unpublished). All these microsatellites can be characterized and brought into usage for pedigree analysis.

CHAPTER 4

ESTIMATION OF HERITABILITY OF HARVEST TRAITS IN SILVER CARP

4.1 Introduction

Heritability is one of the most useful genetic indices about a population, measuring the proportion of the phenotypic variance of a trait that is predictably transmitted to the next generation. Phenotypic variance (V_P) of a character is the combined result of genotypic variance (V_G) and environmental variance (V_E). Genotype-environment correlation ($2Cov_{GE}$) and interaction (V_{GE}) can also contribute to V_P. Genotypic variance arises from three sources: dominance genetic effect (V_D), additive genetic effect (V_A) and epistatic genetic effect (V_I). The total phenotypic variance can thus be shown as the sum of all these contributing components:

$$V_P = V_D + V_A + V_I + V_E + V_{GE} + 2Cov_{GE}$$

Of all these variance components only the V_A of the parental generation is passed on to the offspring generation while all other components are created anew. This fact makes additive genetic variance an important attribute of a population. Additive variance is measured as a proportion of total phenotypic variance and is referred to as heritability.

Heritability is a property of a population and it is not static. The estimate of heritability changes with any changes in size of the population, its gene frequencies and the environmental situation to which the population is subjected. Thus estimates of heritability for the same character on different populations of a species may show a wide range of variation. Heritability varies greatly between traits as well. Generally traits which are closely related to reproductive fitness show the lowest values of heritability as natural selection tends to eliminate less fit alleles and hence variation that produces less fit phenotypes (Kruuk, 2004). On the contrary, characters which are not important determinants of natural fitness show higher values of heritability. In general growth rates show a high level of heritability (Gjedrem 1983).

Knowledge of the heritability of traits is important on several grounds. In wild populations, knowledge of heritability can help understanding of the patterns of evolution and natural selection (Moore and Kukuk 2002, Kruuk 2004). In farmed populations, it has practical implications in predicting the response of commercially important traits to selective breeding programmes and in estimating the breeding values of animals (Falconer and Mackay 1996; Gjedrem and Olesen 2005). Higher heritability indicates a greater potential rate of improvement of a trait through selection.

Heritability of a character is estimated from the degree of resemblance of the phenotype between relatives as this provides a means of estimating the additive genetic variance. The advent of marker technology has tremendously facilitated heritability studies both in natural and farmed populations by allowing analysis of pedigree and relatedness. Since in natural populations, it is often difficult to track the relationships between individuals, previously almost all studies regarding heritability were conducted on farm reared populations. With highly polymorphic molecular markers such as microsatellites, it is now possible to estimate the relatedness (r) of individuals or reconstruct the relationship categories without any pedigree information. Regression of pairwise relatedness estimates on pairwise phenotypic similarity then provides the heritability estimation (Ritland 1996). On the other hand, in farm animals, unlike natural populations, heritability is estimated from carefully structured breeding where pedigree information is maintained. In such breeding programmes molecular markers can be used to assign individuals to the correct relationship categories after they have been reared communally from very early stage of life when they were too small to be physically tagged, such as in fish. Communal rearing of different families is important in heritability studies to reduce the bias from environmental variations.

This chapter focuses on the estimation of heritability of harvest weight and length in silver carp. As has been discussed in Chapter 1, silver carp is one of the most popular carp species in Asian countries and especially in Bangladesh. Heritability was estimated on farmed silver carp as a pre-requisite for planning a selective breeding programme.

The following subsections describe the principle of heritability estimation followed by the practical means of estimation using statistical tools.

4.1.1 Estimation of heritability—principle

Resemblance between any sorts of relatives can be used for the purpose of heritability estimation although the use of some relatives produces biased estimates compared to others. The most frequently used relationships are offspring with parents, and between half-sibs and full-sibs. Covariance of a phenotype between relatives is the basis for estimating the degree of resemblance between relatives and heritability. The covariance differs in amount and proportions according to the sort of relationship. Both genetic and environmental sources of variance contribute to the covariance of relatives.

4.1.1.1 Covariance between relatives and heritability

If epistatic interaction among loci and environmental impact on phenotype are assumed to be absent, the generalised formula of covariance is: $\text{cov} = rV_A + uV_D$. Here, r is the coefficient of additive genetic effect (also called the relatedness co-efficient) and u is the coefficient of dominant genetic effect which represents the probability of the relatives having the same genotype through identity by descent. It is zero unless the related individuals have a path of co-ancestry through both of their respective parents, as have full sibs and double first cousins. The values of ***r*** and ***u*** for different sets of relatives are summarized in Table 4.1.

Table 4.1: Coefficients of variance components in estimating covariance of relatives (after Falconer and Mackay 1996)

Relationship	***r***	***u***	**Covariance**
Monozygotic twins or clones	1	1	$V_A + V_D$
Dizygotic twins or full-sibs	1/2	1/4	$\frac{1}{2}V_A + \frac{1}{4}V_D$
Half-sibs	1/4	0	$\frac{1}{4}V_A$
First cousins	1/8	0	$\frac{1}{8}V_A$
Double first cousins	1/4	1/16	$\frac{1}{4}V_A + \frac{1}{16}V_D$
Parents-offspring	1/2	0	$\frac{1}{2}V_A$
Grandparent-grandchild	1/4	0	$\frac{1}{4}V_A$
Aunt-niece	1/4	0	$\frac{1}{4}V_A$
Uncle-nephew	1/4	0	$\frac{1}{4}V_A$

Epistatic interaction can, however, contribute to the covariance of a trait between relatives, even though the magnitude is generally not very great. Thus, taking this factor into consideration, covariance is expressed as:

$$Cov = rV_A + uV_D + r^2V_{AA} + ruV_{AD} + u^2V_{DD} + r^3V_{AAA} + r^2uV_{AAD} + ru^2V_{ADD} + u^3V_{DDD}$$

Once the covariance component is measured the heritability is estimated as:

$$h^2 = \frac{\frac{1}{r}Cov}{V_P}.$$

The environment can contribute to make relatives resemble each other more when they share a common environment. This can create an upward bias in heritability estimates. The sources of common environmental variance (V_{EC}) are many and varied, and arise from environmental factors such as nutrition, climatic conditions, and maternal effects. Much of the V_{EC} can be eliminated by planning a suitable experimental design. For

example, the members of the same family can be randomized over different rearing environments or members of different families can be reared under the same environment.

4.1.1.2 Choice of relatives and breeding design in heritability experiment

When the heritability is being determined in a wild population experimenters need to use any type of relative that is available. However, when there is opportunity such as when working on a captive population, it is important to choose relatives as well as breeding design to eliminate the effects of non-additive covariance as much as possible. Generally half-sibs are preferred over full-sibs because full-sibs introduce bias due to dominance and common environmental variances. Again paternal half-sibs are preferred over maternal half-sibs or the regression of the offspring on father are preferred over the regression of offspring on mother as the variance component from mothers sometimes gives an over-estimation due to maternal effects (Falconer and Mackay 1996).

While choosing relatives is important for estimation of heritability, so is the breeding design to get the best estimate. Nested designs, where a single male is mated with several females (or vice versa) have been frequently used in livestock breeding for heritability estimation. In the case of fish, high fecundity also allows the use of factorial design (Blanc 2003). The advantage of using a factorial design is that it allows the estimation of both additive and non-additive genetic variances which is not possible with nested designs (Gjerde 2005). The relative efficiency of different types of mating designs in heritability estimation has been studied by a number of workers such as Berg and Henryon (1998), Dupont-Nivet *et al.* (2002) and Blanc (2003) using stochastic and deterministic simulation. These studies showed a superiority of factorial designs over nested designs. Berg and Henryon (1998) observed that a partial factorial design (25

males each mated with 2 females from a lot of 25 females in an overlapping manner; i.e. sire 1 was mated with dams 1 and 2, sire 2 was mated with dams 2 and 3 and so on) gave a more precise estimate than a full factorial design (25 sires x 50 dams). Studies by Dupont-Nivet *et al.* (2002) and Blanc (2003) have shown that a factorial design using a larger number of males and fewer females was the most useful design. Since heritability is generally estimated through males, the greater the number of sires, the more precise the heritability estimate. Dupont-Nivet *et al.* (2002) also studied the optimum family sizes for heritability estimation. Their study shows that accurate estimation of low heritability traits requires a larger family size. The optimum sizes were 3-5 offspring per family for h^2= 0.5, 5-8 offspring per family for h^2= 0.25 and 12-20 offspring per family for h^2= 0.1.

4.1.2 Statistical methods for partitioning variance components

The key points for estimation of heritability are to measure the phenotypic variance of a trait in a population, to partition the variance into observational components, and then to derive the causal components (i.e. additive, dominance genetic variance etc.) from the observational components. Different statistical methods have been devised for estimation of observational components of phenotypic variance. The most frequently used methods are: parent-offspring regression, Analysis of Variance (ANOVA), Maximum Likelihood (ML), and Restricted Maximum Likelihood (REML).

Regression allows us to make inferences about how independent variables determine or predict a dependent variable. This approach is generally applied when heritability is estimated from parent-offspring relationship. The regression coefficient of offspring phenotype on parental phenotype indicates the degree of resemblance between parents and offspring and is used for heritability estimation.

Analysis of variance (ANOVA) has been used most frequently for partitioning of variance components. In general, ANOVA analysis requires that individuals can be assigned to groups with the same degree of relationship for all members such as full-sib or half-sib (Meyer 1989). Genetic influences are inferred on the basis of phenotypic resemblances among relatives compared to lower or lack of resemblance among unrelated individuals. Phenotypic variance is partitioned into within and between family components, both of which can be interpreted in terms of covariances between relatives.

Estimates of variance components by regression and ANOVA, are obtained by least-square statistical analyses. These methods ideally require a balanced design and random sampling of parents from the population for unbiased estimates of the heritability (Falconer and Mackay 1996). Data arising from animal breeding experiments, however, are seldom balanced. Furthermore, individuals used in animal breeding improvement programmes are often selected groups rather than random. To cater with the needs of these situations ML and REML methods have been developed.

Maximum likelihood (ML) methods offer a number of advantages over ANOVA:

- They can account for selection under certain conditions.
- They can handle unbalanced data.
- They can accommodate any structure of genetic relationships in the data such that different types of relationship categories or distant relatives and relatives from several generations can be analysed. This allows utilization of all available information (Falconer and Mackay 1996).

The ML approach functions by calculating the likelihoods of particular values of the parameters (e.g. variance components). Initial values are chosen and an iterative computer algorithm is used to find the combination of parameters at which the

likelihood is the maximum for a given set of data. However, a major drawback of ML estimation is that in mixed model analysis (which analyse both fixed and random effects simultaneously) the fixed effects are treated as if they were known, i.e. the loss in degrees of freedom due to fitting these effects is ignored. If the model of analysis comprises many fixed effects, as is almost invariably the case for animal breeding data, this can yield considerably biased estimates. In particular residual variances can be severely underestimated. ***Restricted Maximum Likelihood (REML)*** is the modified ML procedure that overcomes this problem by incorporating the degrees of freedom for fixed effects. Both ML and REML are computationally intensive. Thus specialized statistical packages (e.g. SAS, GENSTAT) are required for ML or REML based analysis.

4.1.3 Objectives of the study

This study had following two objectives:

1. To study the effect of differential contribution of families on effective population size.
2. Estimation of heritability for harvest weight and length.

4.2 Materials and methods

4.2.1 Fish stock

The broodfish used for the present study came from the broodstock maintained at the NFRDMP (North-West Fisheries Resource Development and Management Project) hatchery at Parbatipur, Bangladesh. This stock of silver carp was established in 1994 by importing Yangtze River silver carp fingerlings directly from China. The stock maintained at NFRDMP consisted of 486 broodfish (Sattar and Das 2002). A system has been developed to maintain the genetic diversity of the stock by maintaining the effective population size (N_e) and by controlling inbreeding. In this system each broodfish is tagged with an electronic PIT tag. For broodstock replacement purposes single pair crosses are used to ensure participation of an equal number of male and female breeders. Using the PIT tag identification it is ensured during the broodstock replacement that each brood fish contributes only once to the next generation. The effective population size is maintained at between 50 and 100 spawners (in 1:1 sex ratio) to keep the inbreeding to an acceptable level.

4.2.2 Breeding design and rearing

Breeding programmes for the present study were performed at the NFRDMP hatchery during the months of August and September 2003. Breeding was performed in three sets, each set consisting of 12 males and 12 females. The three sets of spawnings were arranged sequentially several days apart, so that the same twelve incubators could be used for each set. Milt and eggs from hypophysed males and females were collected by hand stripping. Milt from all 12 males (in a set) was stripped shortly before the females were ready for ovulation and held in separate containers in a refrigerator. As each female was stripped, the eggs were split into 4 batches. Each batch of eggs was fertilized with milt from a single male i.e. eggs from one female were fertilized with

milt from four males. Again milt of a single male was used to fertilise eggs of four females. The breeding design is shown in Figure 4.1. Four equal sub-batches of fertilized eggs from each female were pooled after water hardening in a single incubator. An equal number of hatched fry from each incubator was taken for communal rearing in a single nursery pond. At the fingerling stage a random sample of about 120 fish was taken from each nursery pond and stocked in a single grow-out pond, giving three grow-out ponds in total. With the assumption that all the broodfish contributed equally, it was expected that each sire and each dam would be represented by 10 offspring. Since silver carp are generally grown in polyculture along with other species, the fish in this experiment were stocked in polyculture ponds. Table 4.2 shows the species composition of the polyculture systems. Each pond was 10 decimals (400 m^2) in size and the stocking density of fish was 45 fish/ decimal. The depth of the ponds was approximately 1 m although that varied depending on evaporation rate and water supply. The fishes were reared for 6 months in the polyculture ponds before harvesting. The harvest weight and length of each fish was recorded for heritability analysis.

Once the first set of fry had been removed from the incubators, another series of crosses was produced using the same design but with different males and females. A third set, again using different males and females, was produced after the spawn from the second set had been removed from the incubators. If all the mating combinations used by this design successfully produced offspring then there should be representation from 48 full-sib families, 12 paternal half sib and 12 maternal half-sib families in each set.

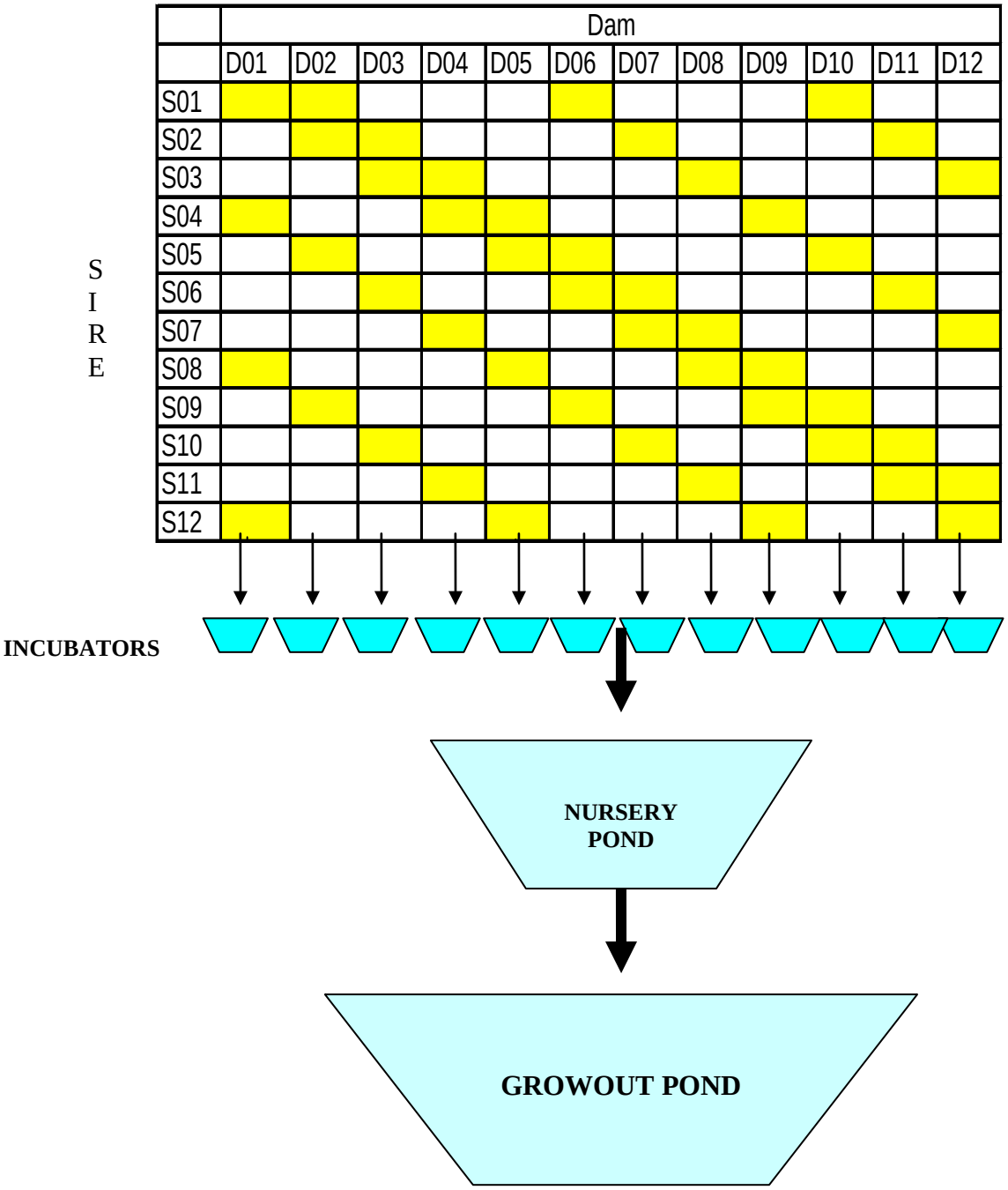

Figure 4.1: Mating and rearing design for estimation of heritability of harvest traits in silver carp (shaded cells indicate matings between these sires and dams)

Table 4.2: Species composition in polyculture system. Pond size 10 decimal (400m^2).

Species stocked	Number of fish	Average weight (g)*	Average length (cm)*
Silver carp	120	5.54	7.62
Catla	30	2.83	6.35
Rohu	60	6.50	<7.62
Mrigal	90	4.39	7.62
Grass carp	30	6.50	<7.62
Silver barb	120	0.79	3.81

*The table shows the average weight and length of fingerlings stocked in pond 1 as a representative of the three trials

4.2.3 Parentage analysis

Since families were reared in communal facilities from a very early stage of life, physical tagging was not possible. Microsatellite markers were used for assigning parents to individual fish after harvesting. Microsatellite analysis was performed on DNA extracted from fin samples that were collected from parents during breeding and from offspring during harvesting. The parentage assignment method and success have already been discussed in the previous chapter (section 3.3.3.2).

4.2.4 Estimation of family contribution

The design of the breeding was such that in each set all parents (both male and female) were expected to be equally represented in the progeny. The number of observed offspring per male and per female was therefore compared to an expected uniform distribution with Pearsons chi-square goodness-of-fit test by GenStat (ver. 8).

To assess the effect of family size, effective population size (N_e) was calculated by the following formula which takes variability in family size into account:

$$N_e = \frac{4(N-2)}{(K_s + \frac{V_s}{K_s}) + (K_d + \frac{V_d}{K_d}) - 2}$$

(Chevassus 1989; taken from Vandeputte *et al.* 2004)

Where N is the total number of offspring, K_s and K_d the mean numbers of offspring per sire and per dam, and V_s and V_d the variances of sire and dam family sizes.

4.2.5 Traits studied

The aim of the study was to estimate the heritability of harvest weight and length of silver carp. The weight and length of individual offspring were recorded during harvesting by the hatchery personnel (it was not possible for me to go to Bangladesh and be present there). Even though the task was performed under expert supervision still there were some doubts if the length measures had all been taken carefully and consistently (hatchery workers are used to take length measures in approximate centimetre values rather than in appropriate fractions). However, it was confirmed that weight data was recorded carefully and accurately.

4.2.6 Partitioning of variance component and heritability calculation

The statistical software GenStat (ver.8) was used for partitioning of variance components. The harvest weight and length data of offspring from all three sets were combined for this analysis. Partitioning of the variance components was carried out by the REML method by fitting a linear mixed model. Sets or ponds, sires and dams were considered as the factors and growth traits i.e. weight and length as variates. The model used for REML analysis was:

$$Y_{rijk} = m + p_r + s_i + d_j + e_{rijk}$$

Y_{rijk} is the length or weight of the k^{th} individual of sire i and dam j in the r^{th} pond. m is the overall constant.

p_r is the effect of pond (or set) r, for r = 1, 2, 3. Effects of spawning date and genetic sampling error are absorbed in the pond effect.

s_i is the effect of sire i, for i =1, 2…… 36; s_i were assumed to be independent and distributed as N(0, σ_S^2).

d_j is the effect of dam j, for j = 1, 2……36; d_j were assumed to be independent and distributed as N(0, σ_D^2).

e_{rijk} is the random error for individual k of sire i and dam j in the r^{th} pond; e_{rijk} were assumed to be independent and distributed as N(0, σ_e^2).

In the above equation, the overall constant (m) and the pond effect (p_r) together constituted the fixed part of the model whereas the sire effect (s_i), dam effect (d_j) and residual error (e_{rijk}) constituted the random part.

The Wald statistic, produced by REML, was used to assess the contributions of the individual terms in the fixed model. To evaluate the contribution of individual terms in the random model, deviance components (-2 log likelihood ratio) of the full model (Set + Sires + Dam) and reduced models (Set + Dam and Set + Sire) were estimated. The deviance due to Sire and Dam was then obtained by deducting the full model deviance from the reduced model deviance. The likelihood ratio test was conducted for significance analysis.

The Genstat programme allows writing commands for advanced functions such as calculation of heritability from the estimate of variance components. The VFUNCTION

directive was used for this purpose. Heritability was estimated in three different ways: (a) from sire variance only, (b) from dam variance only, (c) combined estimate from sire and dam variance. Since sire and dam variances are equivalent to the covariance of paternal and maternal half-sibs respectively, the following equations were used for heritability estimation from sire or dam variance.

$$h_S^2 = \frac{4 \times \sigma_S^2}{\sigma_S^2 + \sigma_D^2 + \sigma_e^2} \quad \text{..…….. for sire half-sib}$$

$$h_D^2 = \frac{4 \times \sigma_D^2}{\sigma_S^2 + \sigma_D^2 + \sigma_e^2} \quad \text{…….for dam half-sib}$$

Where σ^2_S is the sire variance, σ^2_D is the dam variance and the σ^2_e is the variance of the residual term.

For combined sire and dam estimate the following formula was used. This is equivalent to the estimate from full-sib covariance.

$$h_C^2 = \frac{2(\sigma_S^2 + \sigma_D^2)}{\sigma_S^2 + \sigma_D^2 + \sigma_e^2}$$

Phenotypic variance was calculated as $\sigma_P^2 = \sigma_S^2 + \sigma_D^2 + \sigma_e^2$.

A Likelihood Ratio Test (LRT) was performed to test the hypothesis that $\sigma_S^2 = \sigma_D^2$. Testing this is important for two reasons. First, if sire and dam components are not equal, it will indicate that the larger variance might have been affected by non-additive factors such as common environmental effect, maternal effect and/or dominance effect (generally affects the dam variance). Secondly, this will give an indication of whether the values of combined estimation of heritability (h_c^2) should be considered reliable.

When sire and dam variances are equal, the combined estimates of heritability are expected to provide more precise values as they use more of the available information. To test the null hypothesis that $\sigma_S^2 = \sigma_D^2$ by LRT, a manual iterative process was carried out so that initial ratios (called gamma, γ values) of sire and dam components with the residual components can be fixed to be equal. Gamma ratios are expressed by following equations:

For sire, $\gamma_S = \sigma_S^2 / \sigma_e^2$ and for dam, $\gamma_D = \sigma_D^2 / \sigma_e^2$.

If $\sigma_S^2 = \sigma_D^2$ then, $\gamma_S = \gamma_D = \gamma$.

In this case the combined estimate of heritability can be expressed as $h_C^{2\prime} = \frac{4\gamma}{2\gamma + 1}$.

The iteration was carried out by varying the γ values in an arbitrary range where the upper and lower limit of the range was set by calculating the ratios of original sire and dam components with residual component (i.e. the components estimated in REML analysis). The deviance value for each cycle of iteration was obtained and heritability was calculated from the gamma values as shown above. The gamma value associated with the least deviance was considered to be the best fit to the available data. Therefore, the heritability estimate with this gamma ratio should also be the most accurate. The difference between the smallest deviance obtained from the manual iteration and the deviance obtained by REML analysis was tested against chi-square tables to get the probability of the likelihood ratio test statistic. A non-significant difference would indicate that there is not enough evidence to reject the null hypothesis. In short the purpose of manual iteration was to compare the heritability estimation obtained with equal gamma values to the original heritability estimation by REML where best gamma ratios were determined by the method itself.

4.3 Results

4.3.1 Differential family representation

The contributions of different sires and dams were analysed based on the parentage assignment result (described in section 3.3.3.2). The expected number of offspring from all sires and dams in a set was the same and the expected number was calculated by dividing the total number of offspring that could be matched to a single family by the number of sires (or dams). Tables 4.3a, 4.3b and 4.3c show the expected and observed number of offspring for all the sires and dams in different sets and also the contribution of different families. The percent contributions of sires and dams in each set are shown in Figure 4.2. The expected number of offspring per sire/dam varied between sets due to differences in survival rate and to a lesser extent due to success in parentage assignment. In Set A, the sires were represented by 5 to 17 offspring and dams by 0-20 offspring when the expected frequency for both sires and dams was 9.5 offspring. The χ^2 test showed a significant difference between the expected and observed frequencies of offspring for sires in Set A (χ^2 = 22.77, 11 *df*, p = 0.020). The difference between expected and observed frequencies for dams was highly significant (χ^2 = 37.16, 11 *df*, p<0.001). The large distortion of representation of dams was strongly influenced by two dams, D06 and D07, which produced 20 and 0 offspring respectively. When these two dams were removed from the analysis there was no more representation distortion (χ^2 = 16.05, 9 *df*, p= 0.066).

In Set B, the number of offspring for the sires ranged from 3 to 17 and for dams from 1 to 17 while the expected number was 9.25. Similarly to Set A, a significant difference was obtained for sires (χ^2 = 20.14, 11 *df*, p = 0.044) but a highly significant difference was observed for dams (χ^2 = 39.78, 11 *df*, p<0.001) between observed and expected frequencies of offspring. Two dams, D20 and D21 produced only 1 offspring each.

After their exclusion from the analysis, the difference was still significant but only at the 5% level (χ^2 = 17.96, 9 *df*, *p* = 0.036).

In Set C, sires were represented by 1-16 offspring and dams by 2-20 while the expectation was 7.8 offspring per sire and per dam. The observed number of offspring for sires was significant at 5% level (χ^2 = 24.06, 11 *df*, *p* = 0.012) but again dams showed highly significant deviations from the expected value (χ^2 = 43.48, 11 *df*, $p<0.001$). In this set one sire produced only 1 offspring while two dams (D25 and D26) produced only two offspring and two dams (D27 and D29) produced more than or almost double the expected number of offspring (20 and 15 offspring respectively). Taking all the sires and dams from three sets which produced viable offspring the average family size was calculated to be 8.86 ± 4.17 (mean ± SD) for sires and 9.11 ± 5.28 for dams.

Table 4.3a: Number of offspring from different sires, dams and families in Set A (the shaded squares indicate allowed matings)

Sire	Dam												Observed no. of offspring per sire	Expected no. of offspring per sire
	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12		
S01	1	2				5				1			9	9.5
S02		1	7				0				1		9	9.5
S03			0	0				3				2	5	9.5
S04	4			0	0				0				4	9.5
S05		7			1	6				1			15	9.5
S06			1			4	0				0		5	9.5
S07				1			0	2				2	5	9.5
S08	4				0			2	5				11	9.5
S09		3				5			0	0			8	9.5
S10			5				0			1	5		11	9.5
S11				2				4			7	4	17	9.5
S12	2				4				5			4	15	9.5
Observed no. of offspring per dam	11	13	13	3	5	20	0	11	10	3	13	12		
Expected no. of offspring per dam	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5		

Table 4.3b: Number of offspring from different sires, dams and families in Set B (the shaded squares indicate allowed matings)

Sire	Dam												Observed no. of offspring per sire	Expected no. of offspring per sire
	D13	D14	D15	D16	D17	D18	D19	D20	D21	D22	D23	D24		
S13	2	0				1				2			5	9.25
S14		1	1				1				0		3	9.25
S15			1	3				0				3	7	9.25
S16	3			4	1				1				9	9.25
S17		1			4	1				2			8	9.25
S18			2			4	0				2		8	9.25
S19				3			3	0				3	9	9.25
S20	6				1			0	0				7	9.25
S21		12				2			0	3			17	9.25
S22			1				4			6	4		15	9.25
S23				6				1			4	3	14	9.25
S24	6				2				0			1	9	9.25
Observed no. offspring per dam	17	14	5	16	8	8	8	1	1	13	10	10		
Expected no. of offspring per dam	9.25	9.25	9.25	9.25	9.25	9.25	9.25	9.25	9.25	9.25	9.25	9.25		

Table 4.3c: Number of offspring from different sires, dams and families in Set C (the shaded squares indicate allowed matings)

Sire	Dam												Observed no. of offspring per sire	Expected no.of offspring per sire
	D25	D26	D27	D28	D29	D30	D31	D32	D33	D34	D35	D36		
S25	0	2				7				0			9	7.8
S26		0	3				5				1		9	7.8
S27			7	3				0				1	11	7.8
S28	0			0	0				1				1	7.8
S29		0			2	3				2			7	7.8
S30			7			0	0				2		9	7.8
S31				1			3	2				4	10	7.8
S32	0				8			2	6				16	7.8
S33		0				1			2	0			3	7.8
S34			3				2			1	0		6	7.8
S35				2				1			0	0	3	7.8
S36	2				5				2			1	10	7.8
Observed no. of offspring per dam	2	2	20	6	15	11	10	5	11	3	3	6		
Expected no. of offspring per dam	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8		

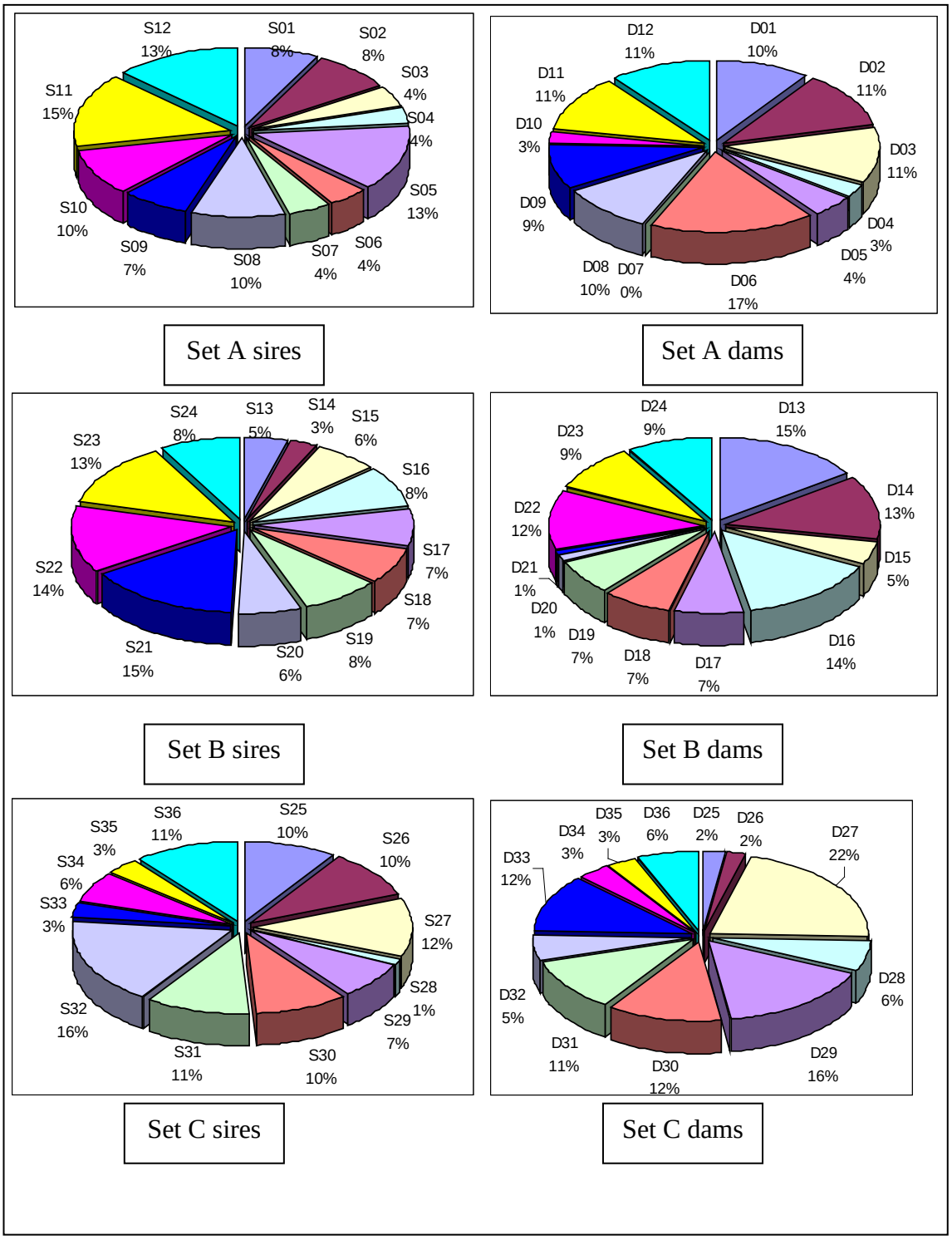

Figure 4.2: Pie charts showing the percent contribution of sires and dams in different sets

Each set of breeding was performed in a way such that 48 full-sib families would be created. However, due to the fact that only a small number of fry/fingerlings (total 360 in three sets) were finally retained for rearing in grow-out ponds (due to the size of the ponds available) along with the variance in family size, none of the sets contained offspring from all the families. In Set A only 36 families, in Set B 39 families and in Set C 32 families produced offspring. Figure 4.3 shows that the biggest 16 families contributed one third of all the offspring of combined three sets, then next biggest 27 families contributed another one third of the offspring and the remaining 64 families contributed the remaining one third of the progeny group.

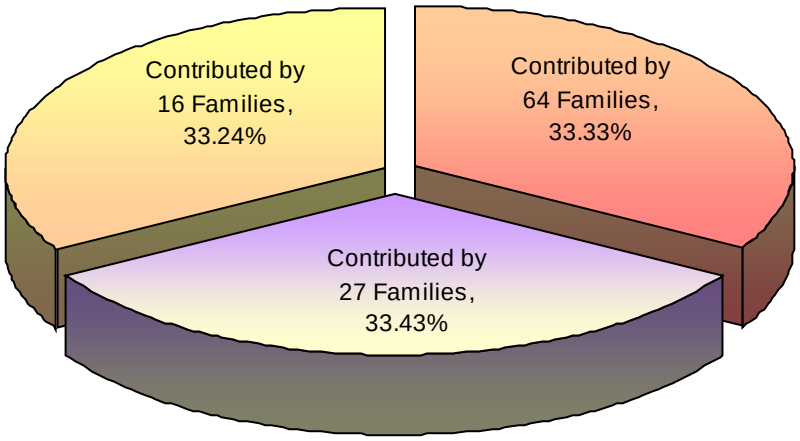

Figure 4.3: Pie chart showing the cumulative contribution of families

If all the parents had contributed to the progeny and the family sizes were equal, the effective breeding number would have been 72. Considering the variability of the family size, the effective breeding number was estimated to be 60.40. This means there was a 16.11% reduction in the actual N_e from the expectation.

4.3.2 Descriptive statistics

In three ponds a total of 360 fish were stocked but due to mortality 331 fish were recovered during harvesting. Overall mortality was 8.1% but the greatest mortality was observed in pond 3 which was as high as 16.7% compared to 5% and 2.5% mortalities in pond 1 and pond 2. Table 4.4 reports the summary statistics of harvest length and weight combining individuals from all three sets which could be assigned to single pair of parents by use of microsatellite markers and hence were used for partitioning of variance analysis. The length and weight traits were found to be highly correlated (r^2 = *0.88*) (Figure 4.4).

Table 4.4: Summary statistics of the growth parameters (3 sets combined)

	Weight (g)	**Length (cm)**
Number of observation	319	319
Mean ± SD	405.3 ± 79.6	33.25 ± 2.06
Minimum	180.0	26.00
Maximum	600.0	38.00
Standard error of mean	4.5	0.12
Variance	6339.1	4.26
Co-efficient of variation	19.6	6.20

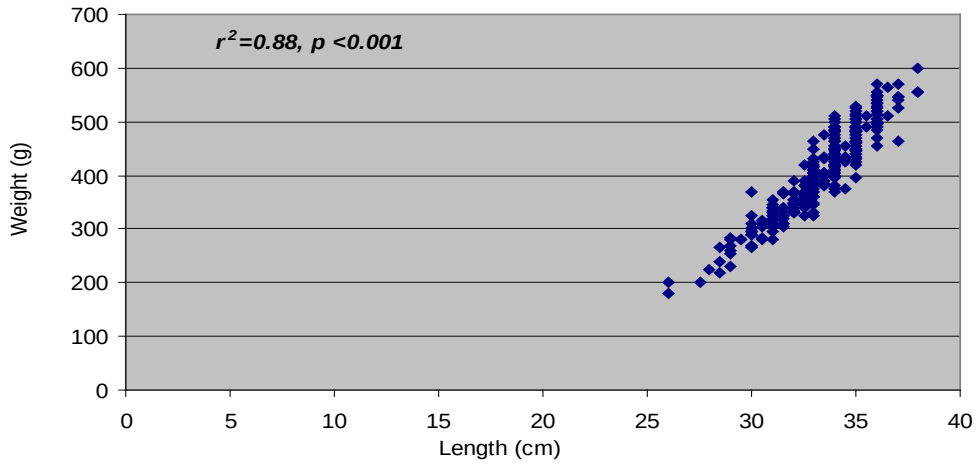

Figure 4.4: Scatter diagram showing the correlation between harvest weight and length of silver carp

4.3.3 Pond (set) effect

The model used in the present experiment for the analysis of variance component considered experimental pond (set) as the fixed effect which included the pooled effect of ponds, spawning date and genetic sampling error. Since in each set there were 24 parents (12 sires and 12 dams), the genetic mean for a set should not deviate considerably from the overall genetic mean by chance. The genetic sampling error is therefore, presumed to be small.

The Wald test (result shown in Table 4.5) indicated that this fixed effect had a significant influence on the harvest parameters. The predicted means of length and weight for the three ponds (sets) are shown in Table 4.6. The best growth parameters were observed in Set A and the lowest parameters were observed in Set C. The grow-out pond where Set C fish were reared suffered from a heavy infestation of filamentous algae. Silver carp is a phytoplankton feeder. It is possible that the heavy growth of filamentous algae might have competed for the limiting nutrients in the pond with

Table 4.5: Result of Wald test for fixed effect in the silver carp heritability experiment

Trait	Wald statistics	d.f	Wald/d.f.	χ^2 probability
Length	87.43	2	43.71	<0.001
Weight	84.48	2	42.24	<0.001

Table 4.6: Predicted means of growth parameters in three sets of experimental fish

	Set A	Set B	Set C
Length			
Predicted mean ± SE	34.66 ± 0.26	33.44 ± 0.26	31.18 ± 0.27
Weight			
Predicted mean ± SE	469.60 ± 10.7	404.50 ± 10.6	329.10 ± 10.9

phytoplankton, decreasing their abundance and in turn reducing the growth of silver carp. However, the effect of other factors, especially the time factor should not be overlooked. The spawning dates between the consecutive sets varied by a week. The Set A breeding was initiated first and the Set C breeding last (a gap of about 10 days between the two sets). The whole breeding and rearing experiment started at the end of August (that is when winter is approaching) when the daylight and temperature were gradually decreasing and hence a gap of 2 weeks might have been important.

Since the silver carp were raised in a polyculture system the growth pattern of other species were also evaluated for comparison between ponds. Other species, however, did not necessarily show the same pattern of growth as silver carp. Even though for most species the pond effect was found to be significant, the ranking of ponds was quite different compared to that for silver carp (Table 4.7). The table show that two species, namely rohu and grass carp showed the best growth in pond C. Both of these species are known to feed on filamentous algae (Masser 2002; Mridula *et al.* 2005). Thus the abundance of filamentous algae in pond C might have favoured the growth of Rohu and grass carp over other species.

Table 4.7: Ranking of the ponds based on the performance of different species

Species	Ranking based on length	Ranking based on weight
Silver carp	A>B>C	A>B>C
Rohu	C>A>B	C>B>A
Thai shorputi	A & C>B	A & C>B
Catla	A>C>B	A>C>B
Grass carp	C>B>A	C>B>A
Mrigal	A>B>C	A>B>C

4.3.4 Genetic parameters

Table 4.8 presents the results of partitioning of variance components by REML analysis and Table 4.9 reports heritabilities of harvest length and harvest weight estimated based on paternal half-sib (sire variance), maternal half-sib (dam variance) and full-sib (combined sire and dam variance) covariance. For harvest weight, heritability estimates ranged from 0.55 to 0.75 and for harvest length, from 0.18 to 0.82. The lowest heritability estimates were obtained from the maternal half-sibs for both length and weight traits while the highest values were obtained when paternal half-sib was used for estimation. For harvest length the three heritability estimates were quite different. The likelihood ratio test was conducted to check if the lowest deviance value estimated from manual iteration considering equal sire and dam variance was significantly different from the deviance value associated with the original estimates of variance components. No significant variation was found either for weight or length data indicating there is not enough evidence to reject the hypothesis that the sire and dam variance is equal. The plotting of re-estimated heritability ($h_C^{2'}$) against the deviance shows that at the minimum value of deviance the heritability estimation was the closest to the original calculation of h_C^2 (Figures 4.5a and 4.5b).

The standard errors of all the heritability estimates ranged from 0.14 to 0.25 for weight and 0.13 to 0.26 for length. The estimates based on combined variance of sire and dam can be considered more accurate as the associated standard errors were the lowest compared to those obtained from the estimates based solely on sire or dam variance.

The estimation of deviance for sire and dam terms and assessing them against χ^2 show that both the sire and dam components have significantly contributed in explaining the deviance of the observed data from expected (Table 4.10).

To find out if there is any dominance variance, sire and dam interaction was also introduced in the random model of REML analysis. For both length and weight data the values were found to be slightly negative (results not shown), indicating negligible dominance variance in the present population.

Table 4.8: Estimates of variance components of silver carp growth traits by REML analysis

	Weight		Length	
Components	**Variance**	**S.E.**	**Variance**	**S.E.**
Sire	587	225	0.472	0.176
Dam	425	202	0.102	0.091
Residual error	2091	183	1.722	0.150
Total phenotypic variance	3103	317.5	2.295	0.223

Table 4.9: Heritability estimates of silver carp growth traits

Calculation method	**Weight**		**Length**	
	h^2	**S.E.**	**h^2**	**S.E.**
Based on paternal half-sib	0.7564	0.2514	0.8222	0.2582
Based on maternal half-sib	0.5479	0.2376	0.1772	0.1557
Based on full-sib	0.6522	0.1358	0.4997	0.1325

Table 4.10: Estimates of deviance of growth traits for sires and dams

Models	**Weight**			**Length**		
	Deviance	**d.f.**	**χ^2 probability**	**Deviance**	**d.f.**	**χ^2 probability**
Set + Sire + Dam	2810.94	313		550.23	313	
Set + Dam	2836.71	314		577.49	314	
Set + Sire	2822.94	314		552.27	314	
Sire[1]	25.77	1	< 0.001	27.26	1	< 0.001
Dam[2]	12	1	< 0.001	2.04	1	< 0.05

[1]Sire deviance and d.f. was obtained by deducting Set + Sire values from Set + Sire + Dam values
[2]Dam deviance and d.f. was calculated by deducting Set + Dam values from Set + Sire + Dam values

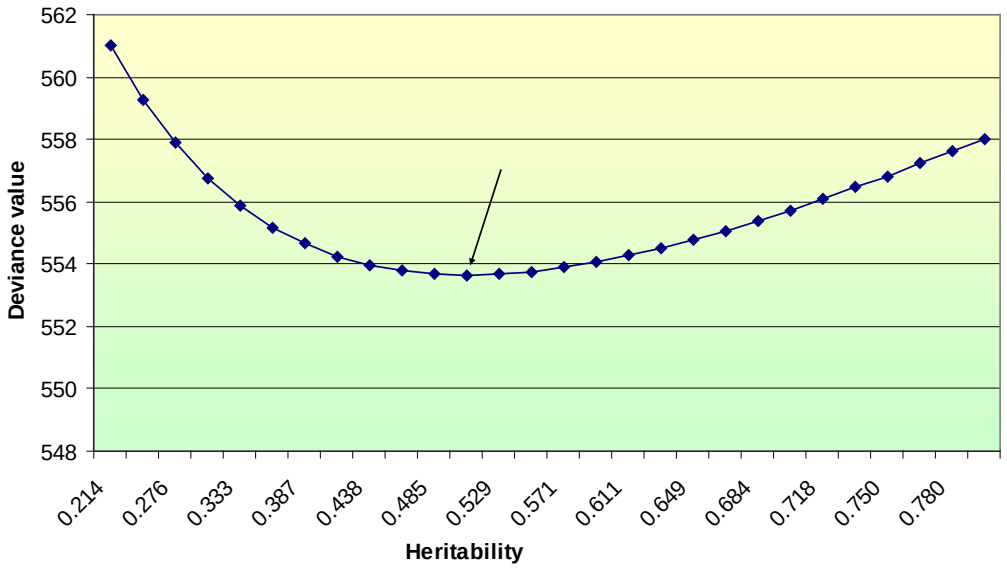

Figure 4.5a: The deviance profile of various heritability estimates for harvest length from manual iteration considering equal sire and dam components. The estimate closest to the original estimate is indicated by the arrow.

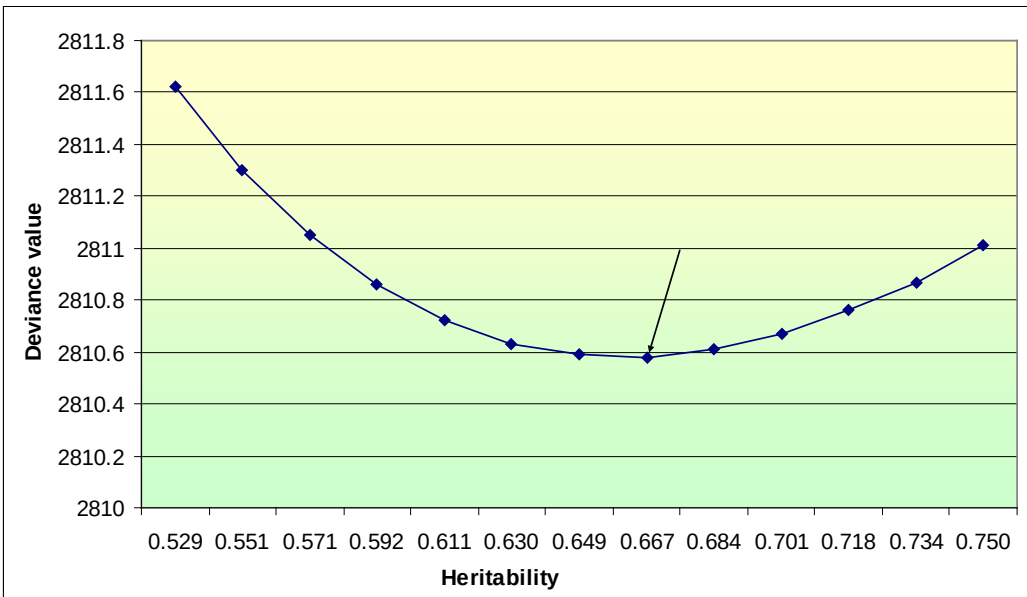

Figure 4.5b: The deviance profile of various heritability estimates for harvest weight from manual iteration considering equal sire and dam components. The estimate closest to the original estimate is indicated by the arrow.

4.4 Discussion

This is the first report on the estimation of heritability of harvest traits of silver carp and also on the use of microsatellite markers in a breeding programme involving silver carp. Heritability estimates can be affected by a number of factors such as breeding design, type of relatives used, number and size of the families, family rearing approach and the method of analysis of variance - all of these factors have been described in the introduction section of this chapter. Heritability in the present experiment was estimated by analysing covariance between half-sib families and between full-sib families which were created in three sets of factorial mating involving 12 sires and 12 dams in each set. All the full-sib and half-sib families were mixed just a few days (6 days) after hatching and were reared in communal nursery ponds and grow-out ponds. REML analysis was applied for partitioning of variance components.

A partial factorial mating design was used in the present experiment instead of full factorial design because this design allowed the use of a greater number of spawners and families compared to what would be possible if a factorial design was used. This approach was found to give a more precise estimate of heritability by Berg and Henryon (1998) compared to a full factorial design. In the current experiment the full-sib and half-sib families were produced in three breeding sets and all the families from a single set were reared in one communal pond (total three ponds for three sets). In designs like this, using a few common sires in each set can provide genetic links between the sets. Therefore, initially it was in the plan to retain two males from the first set for breeding in the second and third sets. This, however, could not be realized because the males were not thought to be able to survive in indoor tanks, and were not able to produce a second batch of sperm in time for the second stripping. Cryopreservation of silver carp

sperm is not yet an established technique that could be exploited to preserve sufficient sperm for breeding at different dates.

In each pond 120 silver carp fingerlings were stocked. Originally, however, the plan was to stock double the number of fingerlings (240) in larger ponds (800 m^2) in order to increase the sample size. However, due to limitations of facilities that was not possible.

4.4.1 Family representation and effective population size

Differential representation of families and spawners in the progeny group is a widely reported phenomenon in fish breeding programmes. Differences in reproductive ability of breeders, fertilization rates of eggs, hatching rates and survival of offspring are the major reasons for unequal contribution of families to next generation. Uneven family distribution and failure of individuals to breed can drastically reduce the effective population size (N_e). For studies regarding the estimation of quantitative parameters such as heritability estimation, highly skewed family size can introduce substantial bias in the estimates even though a sophisticated approach like REML is applied for partitioning of variance components, which can generally handle unbalanced data. Such studies therefore, require careful designing of breeding and rearing phases so that uneven representation of families can be minimized. Techniques such as pairwise mating or controlled mating by stripping of eggs and sperms over mass spawning can considerably alleviate this problem by keeping the sex ratio 1:1. Fishback *et al.* (2002) suggested dividing the eggs collected from each female into equal aliquots (corresponding to the number of males used) and fertilizing each batch of eggs separately with the milt from a single male to minimize the family size differences due to sperm competition. They also suggested incubating fertilized eggs from each female separately so that an equal number of viable progeny can be selected from each female

and mixed for communal rearing in the nursery. In the current experiment a similar approach was applied to control the family size. The result of the present study confirms that this design can be a good approach for reducing the effect of family contribution on N_e. In the current study the loss of N_e was only 16.11%. In a similar study on common carp Vandeputte *et al.* (2004) reported a reduction of 21% in N_e. These reductions in N_e are quite small compared to those reported for other species using mass spawning. For instance, in mass spawning events in gilthead seabream the loss of N_e ranged between 67-73% (Brown *et al.* 2005), in red sea bream (*Pagrus major*) the reduction in N_e was as high as 75% from the actual number of breeders (Perez-Enriquez *et al.* 1999) and in Japanese flounder N_e decreased by 80% in the first generation (Sekino *et al.* 2003). The drastic loss of N_e in all these mass spawning events was mainly due to the total failure of a large number of breeders to produce any offspring at all. Unlike these examples, in the present study only one female out of 72 breeders (36 males and 36 females) failed to contribute any offspring to the next generation.

Even though pooling an equal number of fingerlings from different families before rearing in a communal pond may help reduce the variance in family size considerably, complete homogenization is almost never possible, because of differential survival rates in different families. In the present experiment the males were represented by 1-17 offspring and females by 0 to 20 offspring. The representation of female breeders was more skewed compared to the males. Similar patterns of differential family representation, especially for dam-descended families, have been observed in common carp as well (Vandeputte *et al.* 2004). Herbinger *et al.* (1995) also found large variance in the family size of females in rainbow trout, even though in their experiment all the 10 sires and 10 dams (mated in factorial design) contributed progeny.

4.4.2 Estimates of heritability and non-additive components

High values of heritability were obtained for both harvest weight and harvest length for most estimations in the present study, indicating a high possibility of utilizing the additive variance through selection. Heritability estimations of growth parameters in some other aquaculture species are shown in Table 4.11. The sample on which the heritability was estimated came from a well-managed population where the genetic diversity has been carefully maintained. This is one likely reason for observing high levels of heritability in this population. Heritability of harvest length estimated in this experiment, however, may not be very reliable as there was some question as to whether the length measurements were taken carefully (discussed in section 4.2.5). Estimations of heritability are often associated with large standard errors. Compared to the heritability estimates of growth traits listed in Table 4.11 the standard errors obtained by the present study are intermediate. The use of a relatively small number of individuals (*n=317*; imposed on the study by the available facilities) and unbalanced family size are two plausible causes for obtaining large standard errors. However, considering the small sample size, the standard error is relatively small.

Heritability in the present study was estimated in three ways: using the covariance of the paternal half-sibs (i.e. using sire component), covariance of maternal half-sibs (i.e. using dam component) and covariance of full-sibs (combining sire and dam components). Generally, analyses of maternal half-sibs and full-sibs give biased and inflated estimates of heritability due to maternal effects, non-additive genetic effects and/or common environmental effects which have been reported in many studies such as in rainbow trout (Aulstad *et al.* 1972; Gjerde and Gjedrem 1984; Gjerde and Schaeffer 1989), in Atlantic salmon (Gjerde and Gjedrem 1984), *Penaeus monodon* (Benzie *et al.* 1997).

Table 4.11: Heritability estimates of growth traits in different species

Species	Breeding design (no. of sires & dams)	No. of offspring	Trait	h^2±S.E	Reference
Common carp	Full factorial (10D, 24S)	522	Weight at 8 weeks Length at 8 weeks	0.33 ± 0.08 (h_S^2) 0.33 ± 0.07 (h_S^2)	Vandaputte *et al.* (2004)
Chinook salmon	Single pair mating (15D, 15S)	170	Weight Length	0.20 ± 0.30 0.38 ± 0.41	Mousseau *et al.* (1998)
Atlantic salmon	Nested (217D, 83S)	9488	Weight (ungutted) Length 	0.44 ± 0.11 (h_S^2) 0.54 ± 0.08 (h_D^2) 0.33 ± 0.11 (h_S^2) 0.53 ± 0.07 (h_D^2)	Gjerde and Gjedrem (1984)
Rainbow trout	Nested (108D, 56S)	4485	Weight (ungutted) Length 	0.19 ± 0.11 (h_S^2) 0.39 ± 0.11 (h_D^2) 0.26 ± 0.11 (h_S^2) 0.38 ± 0.09 (h_D^2)	Gjerde and Gjedrem (1984)
Rainbow trout	Full factorial (25 sets each with 2D & 2S)	1669 803 794	Weight at 8 weeks Weight at 1 year Weight at 2 year	0.29 ± 0.15 (h_S^2) 0.20 ± 0.12 (h_S^2) 0.17 ± 0.12 (h_S^2)	McKay *et al.* (1986)
Rainbow trout	Full factorial (48D, 2S)	488	Weight Length	0.55 ± 0.03 (h_D^2) 0.61 ± 0.05 (h_D^2)	Fishback *et al.*(2002)
Rainbow trout	Partial factorial (30D, 30S)	3290	Weight (on 52 days) Weight (on 215 days)	0.00 ± 0.32 0.53 ± 0.27	Henryon *et al.* (2002)
Nile tilapia	Nested (40D, 56S)	1884	Weight Length	0.26 ± 0.12 0.25 ± 0.12	Rutten *et al.* (2005)
Penaeus monodon	Nested (18S, 36D)	360	Weight at 10 weeks Length at 10 weeks 	0.01 ± 0.002 (h_S^2) 0.39 ± 0.004 h_D^2) 0.12 ± 0.07 (h_S^2) 0.30 ± 0.11 (h_D^2)	Benzie *et al.* (1997)
Sea Urchin	Nested (33D, 11S)	1320	Weight at 8 months Weight at 12 months 	0.50 ± 0.07 (h_S^2) 0.57 ± 0.07 (h_D^2) 0.39 ± 0.04 (h_S^2) 0.33 ± 0.05 (h_D^2)	Liu *et al.* (2005)
Silver carp	Partial factorial (12S, 12D) in 3 sets	319	Harvest weight (at 6 months) Harvest length (at 6 months) 	0.76 ± 0.25 (h_S^2) 0.55 ± 0.24 (h_D^2) 0.65 ± 0.14 (h_C^2) 0.82 ± 0.26 (h_S^2) 0.18 ± 0.16 (h_D^2) 0.50 ± 0.13 (h_C^2)	Estimates from present study

Maternal effects can be estimated by comparing the dam component with the corresponding sire component (McKay *et al.* 1986). In contrast to the usual expectation, the dam variances for both harvest weight and length were found to be smaller than the corresponding sire variances in the present experiment, even though the differences were not statistically significant. Maternal effects are generally quite substantial in species where the offspring receives both prenatal and postnatal maternal care such as in mammals. In most cases of fish and shellfish, however, maternal effects are related to variation in egg size and egg quality resulting in differential hatching and survival performances in families during the initial stages of life (Gjedrem and Olesen 2005). The maternal effect has been observed to decrease rapidly with age e.g. in rainbow trout (McKay *et al.* 1986; Henryon *et al.* 2002) and black bream, *Acanthopagrus butcheri* (Doupé and Lymbery 2005). In the present experiment, the variance components were estimated only once, i.e. at harvest. It was, therefore, not possible to evaluate if maternal effects were present during the earlier stages of development and growth in silver carp.

Most other studies reporting a smaller dam component than sire component used hierarchical mating designs with sires nested within dams (rather than more frequently used design of dams nested in sires) (Nenashev 1966; Gall and Huang 1988; Gunnes and Gjedrem 1981; Benzie *et al.* 1997). Vandeputte (2003) suggested that if sires are nested in dams and offspring are initially reared separately, then the sire component estimate might be larger than the dam component. In this case, the sire component contains the common environmental and dominance variance. In the present case, it is difficult to explain the reason for obtaining a smaller dam component than sire component. Since the differences between corresponding sire and dam components were not found to be significant there is no evidence to suggest the presence of non-

additive effects. The lower dam component estimations may have been just a chance factor, perhaps due to small sample size.

The variance component due to interaction between sire and dam represents one-quarter of the dominance genetic variance and any of the environmental variance that is common to all members of a full sib family (McKay *et al.* 1986). In the current experiment the sire x dam interactions for both traits were found to be negligible, indicating the absence of dominance and common environmental effects in full-sib groups. According to Falconer and Mackay (1996), when non-additive effects are absent it is better to consider the combined estimate (using full-sib covariance) since this should be more precise because it utilizes more of the available information. The lower standard error associated with the combined estimate in the present experiment also confirms this fact. The absence of common environmental effects is expected in the present study because all the families were reared together in the same pond from the very beginning, as microsatellite markers were used for retrospective identification of families. Common environmental effects can create significant bias if the families are reared separately at the initial stages and then later mixed. The common environment effect from separate early rearing was estimated by Winkelman and Peterson (1994) in Chinook salmon. The result of their study suggested that the magnitude of the common environmental effect is related to the length of time the members of a full sib family share a common environment. Herbinger *et al.* (1999) compared the early growth performance of Atlantic salmon families reared in separate and communal tanks. Individuals of the same family were reared both in separate and communal tanks and the stocking density in both cases were kept the same. Their result indicated that rearing families in separate tanks resulted in an artificially high level of variation among families originating from environmental differences rather than genetic differences

among families. Calculated heritabilities of length and weight from separate rearing were more than twice the estimates from the communal rearing, indicating probable environmental bias induced by separate rearing.

4.4.3 Conclusions

The purpose of the present study was to assess if the silver carp population being maintained in the NFRDMP hatchery can be improved for harvest traits through a selection programme. The high heritability estimates indicate that the traits should respond favourably to a selective breeding programme. When the heritability of a trait is medium to high, individual or mass selection is superior to family selection. Moreover, mass selection is a cheaper and easier method of selection. Therefore, improvement of the silver carp stock using this approach can be considered. However, mass selection programmes often suffers from rapid accumulation of inbreeding, which can soon counteract the gain achieved through selection programme. The mating and rearing techniques used in the present study was very efficient in maintaining a high level N_e and hence can be emulated in future breeding programmes.

Despite the distinct benefits of using microsatellite markers for pedigree analysis and hence communal rearing of families, their applications have been limited by a number of factors including unavailability of markers for the species to be studied, lack of facilities for microsatellite assay and/or high cost associated with the analysis. Comparative cost-benefit analyses of using microsatellite markers and other options (such as separate rearing of families followed by physical tagging) will have to be carried out to determine the applicability of microsatellite markers in a given context.

CHAPTER 5

MONITORING A SELECTIVE BREEDING PROGRAMME IN COMMON CARP (*Cyprinus carpio*) USING MICROSATELLITE MARKERS

5.1 Introduction

The present experiment is part of a mass selection programme for common carp that is currently running in the state of Karnataka, southern India. Common carp is one of the most widely grown species in this state. Even though it is an exotic species to the region it has become popular because of year-round availability of seed. According to recent statistics approximately 35% of the total inland aquaculture production in Karnataka is common carp (Basavaraju *et al.* 2003). The fish is grown either in monoculture or more commonly in polyculture with Indian major carps. The available stocks of common carp in Karnataka are thought to have originated from two sources: the "German" strain introduced to India in 1939 and the "Bangkok" strain in 1957 (Jhingran 1991). They have been inadvertently mixed over many generations to give the current local stock. The local stocks of common carp are characterized by early sexual maturation and multiple spawning in a single year. Often they become sexually mature at an age of six months and sometimes at weights below 100g (Basavaraju *et al.* 2002). Although early maturation and multiple spawnings allow large scale hatchery production of common carp seed, these features are not desirable when the goal is to grow the fish for market. Early maturation diverts energy from somatic growth to gonad development leading to stunted growth. This has been reported to have significant impact on common carp yield in Karnataka. Other potential causes of slow growth are loss of genetic diversity due to establishment of the broodstock from a very small number of fish, accumulation of inbreeding and inadvertent negative selection - all of which are common features in most of the hatcheries in Karnataka. Several potential solutions to the problems of early maturation and slow growth have been suggested and investigated by a team of workers in Karnataka under a research project based at the University of Agricultural Sciences, Bangalore and funded by DFID-AFGRP, UK. The options investigated were:

- Production of sterile individuals by inducing triploidy;
- Production of a monosex population so that reproduction can be avoided in the absence of opposite sex;
- Evaluation of other strains/stocks of common carp to identify late maturing and faster growing strains/stocks and to develop a selection programme based on this.

In the context of the present study the third option is relevant and therefore is discussed further.

5.1.1 Evaluation of different stocks of common carp

Basavaraju *et al.* (2003) evaluated six stocks of common carp with different origins and histories (Table 5.1). Two of these were local stocks and the other four were introduced from different geographical locations outside India. The L-BRP stock came from the largest state hatchery in Karnataka. The large number of broodfish maintained in the hatchery is thought to have ensured a reasonably high effective population size although little effort had been made towards other genetic management issues. The L-FRS stock had been maintained at the Fisheries Research Station at Hesaraghatta, Bangalore, India for occasional seed production and is thought to have originated from the same source as the L-BRP stock. This stock had not been monitored properly and had a very low effective population size. The Amur and the P3 stocks were procured from the live gene bank of HAKI, Szarvas, Hungary. The Amur stock descended from the wild fish from the River Amur in East Asia and had been maintained at HAKI for several generations. The P3 stock was developed through selective breeding from fish of European origin. The RJ stock came from the Research Institute for Freshwater Fisheries (RIFF), Bogor, Indonesia and is thought to have originated from one or more introduced stocks from Asia and Europe. The SV stock was brought from Vietnam and had been produced by

mass selection from a base population composed of three different stocks (Hungarian, Indonesian and Vietnamese). The evidence suggests that the SV stock has reached a plateau of selection response as in the sixth generation the heritability of harvest weight had declined to nearly zero.

Table 5.1: List of the common carp stocks evaluated for growth performance (from Basavaraju *et al.* 2002)

Full name	Abbreviation	Source	Year obtained
Local stock (FRSH)	L-FRS	Fisheries Research Station, Hesaraghatta, Karnataka	Already present
Local stock (BRP)	L-BRP	BRP state hatchery, Karnataka	2000
Amur	Amur	Research Institute for Fisheries, HAKI, Szarvas, Hungary (originates from river Amur, E.Asia)	2000
P-3	P-3	Research Institute for Fisheries, HAKI, Szarvas, Hungary	2000
Selected Vietnamese	SV	Research Institute for Aquaculture No. 1 (RIA 1), Ha Bac, Vietnam	1998
Rajadhanu	RJ	Research Institute for Freshwater Fisheries (RIFF), Bogor, Indonesia	2001

The stocks were assessed for their growth performance under different environmental and aquaculture conditions, for genetic variation and for some reproduction related traits such as gonado-somatic index (GSI) and dress-out percentage. Among these traits, harvest weight was the most important for commercial purposes. The performances of crosses between different stocks (except L-FRS) were also evaluated and their relative heterosis values were measured. The assessment of the stocks and their crosses were carried out in six trials over a period of three years. In each trial, four or five different groups of fish were stocked into a series of different culture systems: polyculture,

monoculture, and concrete tanks. Both for mono- and polycultures different stocks of common carps were reared in the same pond.

Important observations arising from the evaluation experiments were as follows:

- For harvest weight the stocks ranked as: Amur> P3> RJ>L-BRP>SV>L-FRS (in order of best performance).
- In terms of lowest gonadosomatic index the ranking of stocks was SV>Amur>RJ>L-BRP>L-FRS>P3. The earliest age of first maturation was observed in the two local stocks, L-FRS (21.6 weeks) and L-BRP (22.7 weeks), followed by Amur (24.1 weeks), SV (25.2 weeks), RJ (26.1 weeks) and P3 (27.0 weeks).
- Mean harvest weight for each stock was positively correlated with molecular genetic variation assessed by RAPD marker ($r^2 = 0.703$; $P < 0.05$).
- None of the inter-stock hybrids were found to be superior in performance to the best performing pure Amur stock.
- Relative heterosis values were negatively correlated to both the mid-parent values of the crosses for harvest weight ($r^2 = 0.704$; $P < 0.01$) and to RAPD assessed molecular genetic variation ($r^2 = 0.805$; $P < 0.01$).
- Sex contributed significantly to variance in final weight. Females were on average 14% larger than males and also had significantly lower mean GSI ($P < 0.05$) values. Dress-out value was also higher in females although the difference was not statistically significantly.

The presence of variation in the stocks for growth performance, age of maturity and GSI suggested the potential of utilizing selective breeding approaches as a means of developing a better strain of common carp in Karnataka. The mass selection approach

was chosen to be the most cost effective and logistically simplest method of selective breeding.

5.1.2 Mass selection vs. other selective breeding approaches

The aim of selective breeding is to identify and select as parents the individuals whose progeny, as a group, have the highest possible additive genetic merit for the trait(s) in question (Fjalestad 2005). The basic effect of selection is an alteration of the population mean. The success of an artificial selection programme is measured by *response to selection* (*R*) which is the change in the mean phenotypic values between the progeny of the selected parents and the whole of the parental generation before selection. The average superiority of the selected parents is called the *selection differential* (S). The relationship between R and S is shown as: $R = h^2S$. Selection differential is often expressed in terms of the phenotypic standard deviation (σ_p) and is called *intensity of selection (i).* It is expressed as: $i=S/\sigma_p$.

Different selection approaches have been developed towards the improvement of farm animals. The important ones include mass selection, family selection, within-family selection, and progeny testing (discussed in details by Tave 1993, Falconer and Mackay 1996, Fjalestad 2005).

Selection based on an individual's own performance is called ***individual or mass selection.*** This is the simplest of all the selection approaches and has the potential to yield the most rapid response. Mass selection, however, cannot achieve the desired success under certain conditions. Firstly, mass selection is not efficient if the traits show low heritability. Secondly, it can only be applied for traits that can be measured on individual live animals and hence is not suitable for traits like carcass weight and flesh quality. Third, mass selection is also not suitable when age differences or other

uncontrolled environmental effects contribute a large portion of the phenotypic variance. It is, therefore, vitally important that the environmental influence is kept the same for all individuals by rearing them under the same conditions and selection is carried out from individuals of the same age group. Apart from these, another major practical drawback of mass selection is that inbreeding can accumulate rapidly which ultimately counteract the improvement achieved though selection process. Small size of the selected population, random breeding among selected individuals without knowledge of their family relationship and unequal family representation are major reasons for high rates of inbreeding in mass selection.

For traits showing low heritability, ***between-family selection*** is preferred over mass selection. In this approach, whole families are either selected or rejected as units. Family selection also overcomes some other limitations of mass selection. For instance, this can be applied to traits that cannot be measured on live animals. This is possible because breeding values can be estimated based on phenotypic observations on a subsample of full-sibs or half-sibs. Moreover, phenotypic variance due to age difference does not affect the selection response. Family selection, however, can be severely impaired when common environmental factors create large environmental variation among the families. The efficacy of family selection also depends on the size of the family. The larger the family size, the closer is the correspondence between the mean phenotypic value and mean genotypic value. The practical difficulties of family selection thus arise from the necessity of maintaining a large number of families and large size for each family to keep inbreeding level under control and to achieve reasonably high selection intensity. This involves more expense and space.

Within-family selection involves choosing the best performing individuals from each family. This selection approach offers a greater advantage over others when common environmental factors contribute a considerable portion of the phenotypic variance. In general, however, within family selection has low efficiency compared to other selection methods (Fjalested 2005).

Progeny testing is another method of selection where individuals are selected on the basis of the performance (or mean phenotypic value) of their offspring. In theory, this approach should be the most ideal one as it allows selecting individuals based on their observed breeding values. In practice, however, progeny testing can be difficult to apply because the selection of the parents cannot be carried out until the offspring have been measured. This limitation makes progeny testing impossible or difficult in species which either spawn only once or suffer high mortality during or after spawning e.g. salmonids.

Combined selection, where more than one method of selection is incorporated in a breeding programme, is in principle the best method. For instance, family and individual selection can be combined where the best individuals from the best families are selected. A selection index is an efficient method to combine information from an individual and its relatives as well as from several traits (Fjalestad 2005).

A mass selection approach was chosen for common carp due to limitations of resources. In order to overcome the potential shortcomings of mass selection such as inbreeding counteracting the effects of selection and heterogeneity due to non-genetic or maternal effects (e.g. hatching date or egg size variation) leading to wrong selection, features of breeding designs suggested by Chevassus *et al.* (1992) and Vandeputte *et al.* (2002)

were followed. The following features were incorporated in the selective breeding programme for common carp:

- Use of a heterogeneous base population which was synthesized by mixing different stocks with greater representations from better performing stocks.
- Maximizing the number of breeders and homogenizing their contributions to the succeeding generation as far as possible. This was attempted by spawning many single pairs of breeders in separate hapas (rather than mass spawning where the contribution of each fish in unknown) and taking equal numbers of fry from each hapa.
- Removal of potential variation due to different spawning date by growing fish for selection in a series of cohorts, each with the same spawning date.
- Utilization of different locations to grow replicate groups from each cohort for selection to include a range of environments in selection programme. Replicate groups also served as a backup against potential losses.
- Maintaining two selected lines and sending males of one line with the females of another line as broodstock to the seed production hatcheries. This allows re-establishment of variability within populations and neutralizes accumulated inbreeding (Pante *et al.* 2001)

5.1.3 Objectives of the study

The focus of the present study was to employ molecular techniques to monitor the changes in different aspects of the common carp selective breeding programme underway in Karnataka, India. Microsatellite markers were used to study the following features:

1. Relative contributions of stocks at different stages of the selection programme
2. Family representation in the selected populations and its impact on effective population size.
3. Loss of genetic diversity at the microsatellite loci.

5.2 Materials and Methods

5.2.1 Design of the selective breeding programme

The base population (F_0) was created by randomly taking a total of 300 fish from the six stocks (discussed above). The representation of the stocks in the base population was kept as: 25% Amur stock, 25% P3, 20% RJ, 20% L-BRP, 5% SV and 5% L-FRS. The relative contributions of the stocks were decided based on the results of stock evaluation trials discussed in section 5.1.1. All the F_0 fishes were PIT tagged and maintained at the Fisheries Research Station, Hesaraghatta (FRSH), Karnataka. Figure 5.1 shows the design of the selection programme diagrammatically.

For spawning, 100 pairs of broodfish were chosen so that the stock representation in the breeder group matched the F_0 stock composition. Breeders within each stock were chosen randomly. Breeding was carried out by randomly placing pairs of male and female fish in separate hapas. Due to random selection of breeding pairs, the progeny group (F_1) consisted mainly of inter-stock hybrids but also some pure bred offspring. Even though 100 pairs of fishes were taken for breeding, eventually, only 55 pairs spawned.

Breeding was carried out on two separate days and the fish spawned in the same day were considered a cohort. Cohort I consisted of the offspring of 28 pairs and Cohort II of the offspring of 27 pairs of broodfish. The spawn of these two cohorts were reared separately until they were ready to be stocked in grow-out ponds in polyculture. Two lines (Line A and Line B) were created from F_1 fingerlings from both cohort I and cohort II. Since all the fishes in a line could not be placed in a single pond the fishes were divided and stocked in a number of grow-out ponds at different locations.

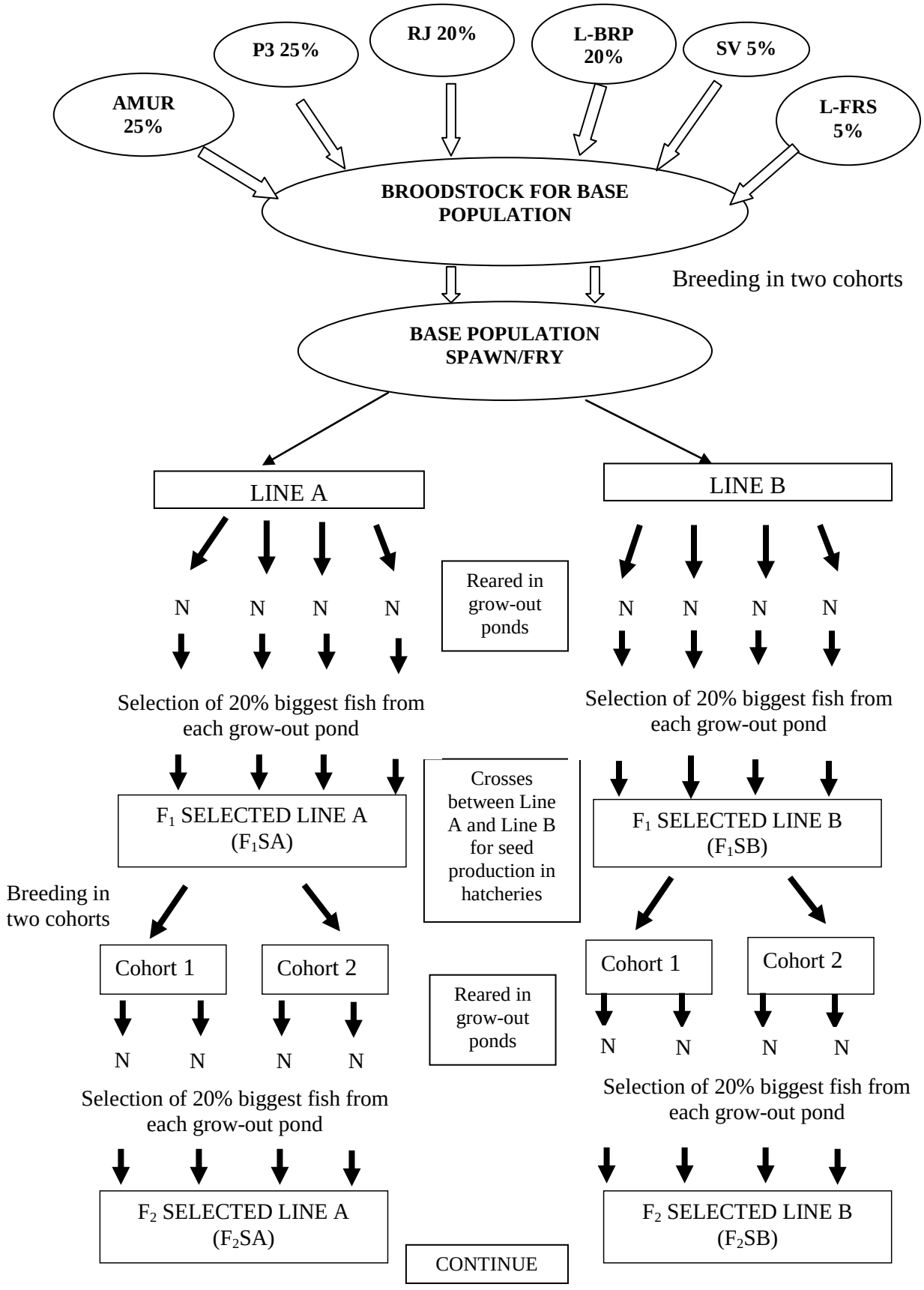

Figure 5.1: Design followed in the selective breeding of common carp

At the end of the grow-out period (six months) each individual fish was weighed and measured and the biggest 20% of the fish in each pond were selected as potential broodfish for producing the next generation (F_2). The target was to select 300 fish from each line. However, from Line A, only 277 fish and from Line B, only 201 fish could be selected. This was due to unexpectedly high mortalities in some ponds. Sex ratios among the selected fish were 193: 84 (females: males) in Line A and 149:52 in Line B, as females grew larger than males. With full awareness of the existence of size dimorphism in different sexes of common carp, the selection was carried out irrespective of sex, because a lower proportion of female fish are ultimately found to be ready for breeding on any date. Weighted mean selection intensities for lines A and B were +1.39 and +1.44 SD units respectively. The selected fish were moved to FRSH from the grow-out ponds, pooled and later PIT-tagged. Fin samples were collected from each fish and fixed in ethanol for future use in parentage analysis.

To create the F_2 generation, the target was to choose minimum 50 pairs of broodfish from each line of F_1 selected fish. In Line A, 75 pairs were ultimately bred. Breeding in Line A was performed in March/April 2005 and Line B was to be bred approximately 6 months later. Only the Line A data is presented here as the information about Line B was not available by the time of writing of the thesis. Breeding in Line A was carried out in two cohorts and the F_2 fishes were reared in several grow-out ponds in the same way as was done for F_1 generation.

5.2.2 Parentage assignment

Parentage assignment of the selected fish was carried out with five microsatellites, namely MFW12, MFW16, MFW18, MFW29 and MFW 11. The markers were chosen from a set of 12 microsatellites that were initially tested for ease and reliability of

parentage assignment. The primer information for the markers was obtained from Crooijmans *et al.* (1997). The chosen loci were highly polymorphic and gave clean and reproducible genotypes whereas other loci suffered from severe stutter bands, amplification of non-specific products or split peaks making the size scoring difficult. The genotyping at these problematic loci did not improve much even after increasing the annealing temperature or decreasing the $MgCl_2$ concentration (as suggested by Mark Vandeputte's team from INRA Fish Genetic Laboratory, France who had used these common carp microsatellites).

Prediction analysis by the programme FAP showed that with only four markers (MFW12, MFW16, MFW18 and MFW29) 98.05% of the individuals could be assigned to a single family. These four markers could be accommodated in two duplex PCRs (described below). Thus initially parentage analysis was carried out with only these four loci and MFW11 was used only for those fishes that could not be assigned with the four loci. FAP was also used for actual parentage assignment. Mismatches of up to 3 alleles were tolerated during parentage analysis. However, in cases of allelic mismatches the offspring and parental genotypes were carefully rechecked to detect any possible typing error and/or presence of null alleles (see section 3.3.3.3 of Chapter 3 for the method of detecting null alleles). The programme CERVUS was used to calculate allele frequency, the PIC values of alleles, the expected and observed heterozygosity (H_e and H_o) and null allele frequency at individual microsatellite locus.

The two duplex PCRs optimized for amplification of four markers (except MFW11) are detailed in Table 5.2 (for the optimization process, see section 2.5 of Chapter 2). The multiplex PCRs were performed in 15 µl volume with the following reagent composition: 2X buffer II, 280 µM of each dNTP, 2 mM $MgCl_2$, 2 U Taq DNA

polymerase, 1 µl of Chelex extracted DNA, primers (concentration specified in Table 5.2) and water to make 15 µl volume. Although for most samples the multiplex genotypes were very clear to score, in some cases using multiplex B, spurious peaks interfered with the size calling. For those individuals genotyping was repeated by singleplex PCR.

Table 5.2: Multiplex PCRs used for parentage analysis in common carp.

	Loci used	**Size range (bp)**	**Attached dye**	**Concentration of primers**	**PCR condition**	
Multiplex A	MFW12	96-126	TET (Green)	0.12 pmol/µl (each primer)	1 cycle	95°C: 5 min
					29 cycles	95 °C: 30 sec
						51 °C: 30 sec
	MFW29	146-224	FAM (Blue)	0.06 pmol/µl (each primer)		72 °C: 90 sec
					1 cycle	72 °C: 25 min
Multiplex B	MFW16	138-220	TET (Green)	0.08 pmol/µl (each primer)	1 cycle	95°C: 5 min
					30 cycles	95 °C: 30 sec
						55 °C: 30 sec
	MFW18	117-179	FAM (Blue)	0.05 pmol/µl (each primer)		72 °C: 100 sec
					1 cycle	72 °C: 25 min

5.2.3 Contribution of stocks and families

Parentage analysis allowed the identification of family and stock from which a selected individual had originated. Chi square goodness-of-fit tests were performed to check the significance of deviation of the observed frequency from expected frequency of stock and family representations at different stages of the selection programme.

5.2.4 Estimation of effective population size (N_e)

Effective population size was calculated using the same method used for N_e calculation for silver carp in Chapter 4 (section 4.2.4).

5.2.5 Assessing the genetic diversity at different stages

The genetic diversity at different stages of the selection programme was assessed by changes in different parameters of the microsatellite loci describing allele profiles and locus informativeness. The parameters studied were: number of alleles, effective number of alleles, polymorphic information content (PIC) and heterozygosity. The estimation of all these parameters has been discussed in Chapter 2 (section 2.8). To measure the proportion of genetic variation remaining in the reared selected population compared to the founding population, "allelic diversity" (A_d) was estimated based on the number of alleles retained at a polymorphic locus (Allendorf and Ryman, 1987). This value was calculated using the formula: $A_d = \frac{n^{/} - 1}{n - 1}$.

Where n is the initial number of alleles present in the base population and $n^{/}$ is the number of alleles retained in the selected populations. Allelic diversity ranges from 1, where all alleles are retained, to 0 where all alleles but one are lost (Allendorf and Ryman, 1987).

The genic differentiation (differentiation in allelic distribution) and genotypic differentiation (differentiation between genotype distributions) between different stages were compared and tested by the programme GENEPOP 3.4 (Raymond and Rousset 1995). The programme estimates the *P* value for each comparison by a Markov chain method. The default settings of 1000 dememorization number, 100 batches and 1000 iterations per batch were used for the *P* estimation by the programme.

5.3 Results

5.3.1 Microsatellite based parentage assignment

Table 5.3 presents information on polymorphism and informativeness of individual loci used for parentage analysis. All the loci were highly polymorphic with 15 to 20 alleles. Heterozygosity values were also very high, ranging from 0.873 to 0.92. High polymorphic information content (PIC) values for the loci (0.857-0.910) indicate high informativeness of the markers. However, one alarming feature was the high positive estimated null frequency associated with all the loci and none of the loci was in Hardy-Weinberg equilibrium. While high null frequency and deviation from HWE indicate the possibility of the presence of null alleles, these may result from other causes as well. If a "population" consists of different subpopulations, deviation from HWE is expected and positive null frequency can be observed. Since the common carp base population used for selective breeding (in which the markers have been characterised) was synthesized from a number of stocks, it is quite expected to observe deviation from HWE. Therefore, parentage analysis was performed with these markers, despite the high null allele prediction. The results of parentage analysis for two lines of F_1 selected fish (F_1S LineA and LineB) are summarised in Table 5.4. Although in Line A the number of selected fish was originally 277, only 236 were available for analysis. Four microsatellites, namely MFW12, MFW29, MFW16 and MFW18 were first applied to assign parentage as they could be accommodated in two multiplex PCR reactions. Although the prediction for single family assignment with 4 loci was 98.05%, a much lower actual assignment (89.7%) was achieved. With the use of the 5^{th} locus (MFW11) on unassigned offspring, the rate of assignment rose to 95.4%, still considerably lower than the expected success.

Table 5.3: Genetic characteristics of common carp microsatellites used in parentage analysis

Locus	A	n	H_o	H_e	PIC	HWE test	Null allele frequency
MFW11	18	109	0.734	0.919	0.908	***	+0.1117
MFW12	15	110	0.845	0.901	0.889	***	+ 0.0292
MFW16	20	107	0.729	0.905	0.894	***	+ 0.1050
MFW18	18	110	0.755	0.873	0.857	***	+ 0.0712
MFW29	20	110	0.727	0.920	0.910	***	+ 0.1161
Mean	18.2		0.758	0.904	0.892		

A= number of alleles
n = sample size
H_o= observed heterozygosity
H_e= expected heterozygosity
PIC = polymorphic information content
*** P<0.001
Null frequency: see section 2.8.5 for explanation

Table 5.4: Summary of parentage analysis

	F_1S Line A	F_1S Line B	Line A & B combined
Number of fish genotyped	236	201	437
Single family match with 4 loci	208	184	392
Percent match with 4 loci	88.16%	91.5%	89.7%
Single family match after using 5th locus	**220**	**197**	**417**
Percent match with 5 loci	**93.2%**	**98 %**	**95.4%**

Null alleles were a major problem in the assignment process. The total assignment stated in Table 5.4 was obtained after manual correction for null alleles. Null alleles were detected in all the loci but the preponderance were in MFW16. At least 13 parents showed null alleles in locus MFW16 while 8 parents in MFW18, 4 parents in MFW29 and 3 parents in MFW12 showed null alleles. It is worth noting that although the predicted null allele frequency for MFW29 was much higher than for MFW16 and MFW18, fewer individuals were affected by null alleles at this locus in actual cases. This confirms contribution of other factors in giving the high null allele prediction. To help explain the method of detecting null alleles, genotypes of a number of offspring from one of the families are presented in Table 5.5, where the sire has a null allele at

locus MHW16. The genotypes of the affected parent and offspring are shown in the table by shaded cells. The genotype of sire at MFW16 is 161/161. The genotypes of the putative null allele affected offspring match with the genotype of the sire (as well as of the dam) at all loci except at MFW16. These offspring failed to show any 161 allele that should have been transmitted by their father.

Table 5.5: Genotype at four microsatellite loci of a sample of offspring from a family showing null allele at locus MFW16 (Shaded cells indicate presence of null allele)

	MFW29		MFW12		MFW16		MFW18	
	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
D45	165	224	101	103	117	135	159	193
S45	163	181	127	127	161	161	138	163
offspring								
AF25	181	224	101	127	135	161	159	163
AF69	163	165	103	127	135	135	138	159
AF90	163	165	101	127	117	117	138	193
AF97	163	165	101	127	117	161	163	193
AF106	163	224	101	127	135	161	163	193
AF120	181	224	101	127	117	117	138	159
AF188	163	224	103	127	135	135	159	163
AM17	163	165	101	127	117	161	159	163
AM41	165	181	101	127	117	117	138	193
AM83	165	181	103	127	117	161	138	159
BF01	163	165	101	127	117	117	159	163
BF33	181	224	101	127	135	135	159	163
BF38	163	224	101	127	117	117	163	193
BF65	163	165	103	127	117	161	163	193
BF74	181	224	103	127	135	161	163	193
BF87	165	181	101	127	135	135	159	163
BF103	165	181	101	101	117	161	159	159
BF107	165	181	101	127	117	117	163	193
BM01	163	165	101	127	135	161	163	193
BM16	181	224	101	127	117	117	163	193
BM17	165	181	101	127	135	161	159	163
BM49	173	224	101	127	117	117	138	193

5.3.2 Representation of stocks at different stages

One of the objectives of the present study was to monitor if the representation of the six stocks in the base population was maintained in different stages of the selection process. Stock composition at three stages was monitored and compared: (i) base population breeders (F_0B), (ii) 1st generation of selected fish (F_1S Line A and Line B separately and combined), and (iii) breeders from Line A of selected fish (F_1SB). Breeding in Line B was still to be performed at the end of this study. Hence only Line A breeders could be analysed. One hundred pairs of broodfish from F_0 population were chosen in a way that the stock composition of the whole base population is perfectly reflected among the breeders. Unfortunately however, only 55 pairs (out of 100 pairs) ultimately reproduced in two cohorts and the stock representation among the breeders (F_0B) became very different compared to the original F_0 composition (shown in Table 5.6). PIT tag identity of individual breeders allowed identification of the stock they originated from. Three stocks, namely, P3, L-FRS and SV were close to their expected representation in F_0B. Amur was severely underrepresented, showing less than half of its expected representation (only about 11% against the expected 25%) and RJ was overrepresented (approximately 28% against the expected 20%). The observed representation of stocks in F_0B was significantly different (χ^2 =13.727, df = 5, P = 0.017) from the original F_0 composition.

Table 5.6: Comparative stock composition in F_0 population and F_0 breeder group

	Amur	P3	RJ	L-BRP	SV	L-FRS
F_0 population	25%	25%	20%	20%	5%	5%
F_0 breeders (F_0B)	10.91%	26.36%	28.18%	24.55%	4.55%	5.45%

For the F_1 selected (F_1S) group, stock representation was calculated from the parentage assignment information. Most of the individuals in this stage originated from inter-stock hybridization. Thus an equal proportion of a hybrid individual was allocated to the corresponding two stocks from where its parents had originated. The fractional contributions from individuals were then summed up for each stock and percent representation of the stocks was calculated. The same procedure was followed for breeders from the Line A selected fish (F_1SB) group as well. The representations of stocks at four stages beginning from the base population through to the F_1SB are compared in Figure 5.2. The figure shows a drastic drop in the Amur representation in the F_1S group (5.28%) from its representation in the F_0B group (10.91%) but a large increase in the RJ stock in the F_1S (42.45%) compared to the F_0B (28.18%). The Amur representation dropped even further by the F_1SB stage (4.66%). The P3 representation was also reduced in the F_1S group (21.82% from 26.36% in F_0B) but again increased in the F_1SB (25.85%) to match the expected contribution. The L-BRP representation dropped from 24.55% in the F_0B group to 20.86% in the F_1S and continued to drop in the F_1SB group (19.91%). The SV and L-FRS stocks more or less maintained their intended contribution across the different stages. The representation of stocks in the F_1S population (combined Line A and Line B) deviated highly significantly from that in F_0B population ($\chi^2 = 48.408$, df = 5, $P < 0.001$).

Figure 5.2 gives an overall picture of the changes of stock representation at different stages of the selective breeding programme. For this figure Line A and Line B of selected fish were combined to estimate the stock representation of whole F_1S group. However, since Line A and Line B will be maintained separately, it is quite likely that after several generations they will diverge considerably from each other. It is therefore important to track the stock composition of individual lines separately. The composition

of Line A and Line B in terms of stock contribution is presented in Figure 5.3 against the expected contribution. The expected contribution of stocks in both the lines was the stock representation observed at the previous stage i.e. in F_0B. The figure shows that in both lines RJ stock dominated all other stocks. In Line A the stocks in terms of their contribution ranked as: RJ (44.09%)> P3 (22.50%)> L-BRP (17.95%)>Amur (5.91%)> SV (5.23%)>L-FRS (4.32%). In Line B the ranking was: RJ (40.61%)> L-BRP (24.11%)> P3 (21.07%)>L-FRS (5.08%)>Amur and SV (4.57% each). The stock compositions of both the selected lines varied significantly from the composition of F_0B ($\chi^2 = 30.695$, df = 5, P <0.001 for Line A and $\chi^2 = 20.221$, df = 5, P <0.01 for Line B).

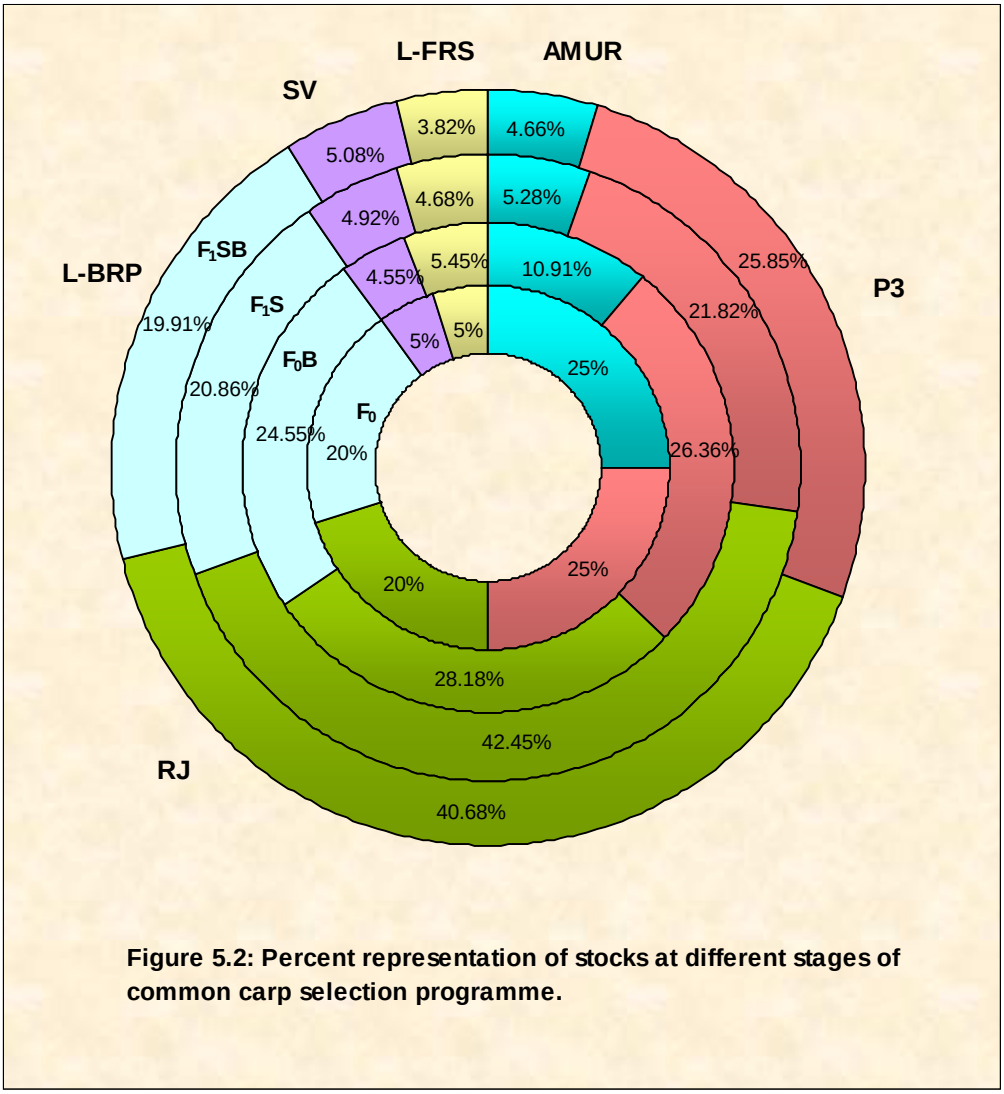

Figure 5.2: Percent representation of stocks at different stages of common carp selection programme.

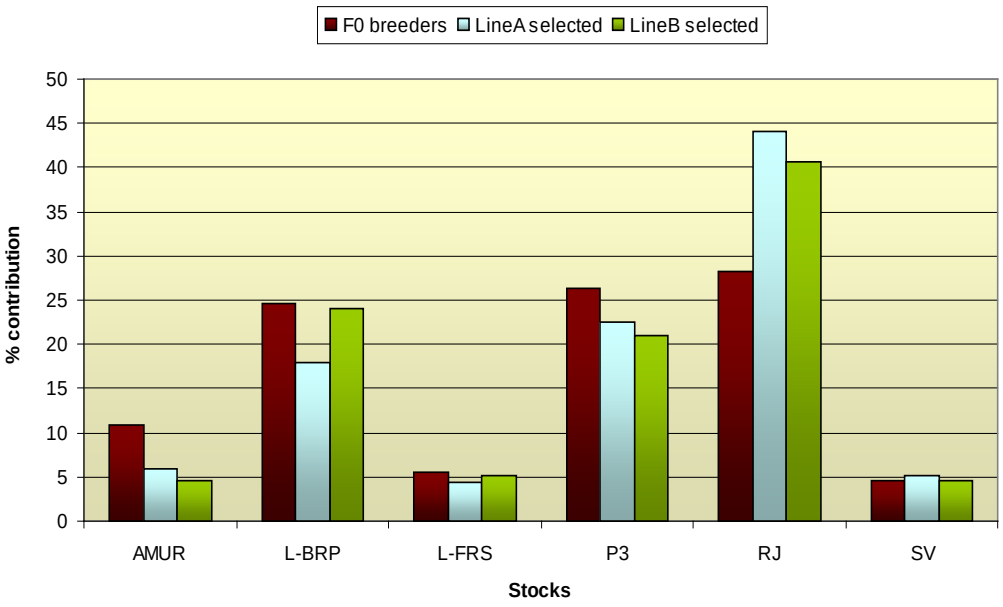

Figure 5.3: Stock composition of Line A and Line B selected fish compared to that of base population breeders

5.3.3 Family contribution and effective population size

Fifty-five pairs of breeders from the base population contributed progeny to the next generation and 20% of their offspring were selected. The contribution of the 55 potential families has been studied based on microsatellite based parentage assignment. This section discusses how families were represented in the F_1 selected population (Line A and Line B separately and combined) and also among the breeders of Line A selected fish which were used to produce the F_2 generation.

Table 5.7 summarizes family representation in the F_1 selected group (F_1S) and in the F_1 breeders (F_1SB) and the impact of differential family representation on effective size of the population (N_e). Out of the 55 families produced, 53 families were represented in the entire selected population. When the two lines were considered separately, the number of represented families was lower than in the combined population. In Line A, progeny contribution from 48 families and in Line B progeny from 46 families were observed.

The appearance of a smaller number of families in individual lines is obviously the result of smaller population size. Family size was highly variable, ranging from 1 offspring to 53 offspring per family when the entire selected population was considered. The average size of the family in the F_1S group was 7.87 ± 8.9 (mean ± SD). Figure 5.4 shows percent contribution of individual families in Line A and Line B and Figure 5.5 presents the distribution of families based on their percent contribution in the entire selected population. Out of 55 families, 37 families were each represented by less than 2% of the offspring (combined contribution 30%), 14 families were represented by 2% to 4% offspring (total contribution about 39%) and only 4 families contributed more than 4% offspring (total contribution 31%). The largest family alone contributed approximately 13% of the entire F_1S population. The observed family contribution was found to be highly significant compared to expected contribution (χ^2=392.07, df = 54, *P* <0.001).

Table 5.7: Summary result of family representation in F_1 selected groups and F_1 breeders group

	F_1 selected fish (F_1S)			**F_1 breeders (F_1 SB)**
	Line A	**Line B**	**Total**	**Line A**
Total number of fish	236	201	437	150
Total number of fish analysed	220	197	417	118
Expected number of families	55	55	55	48
Observed number of families	48	46	53	43
No. of offspring per family	1-33	1-20	1-53	1-11
Effective population size (N_e)	46.14	54.41	49.03	57.21
Census population size* (N)	110	110	110	96
N_e/N	0.42	0.49	0.45	0.60

*Census population size is the total number of male and female parents as the sex ratio was 1:1.

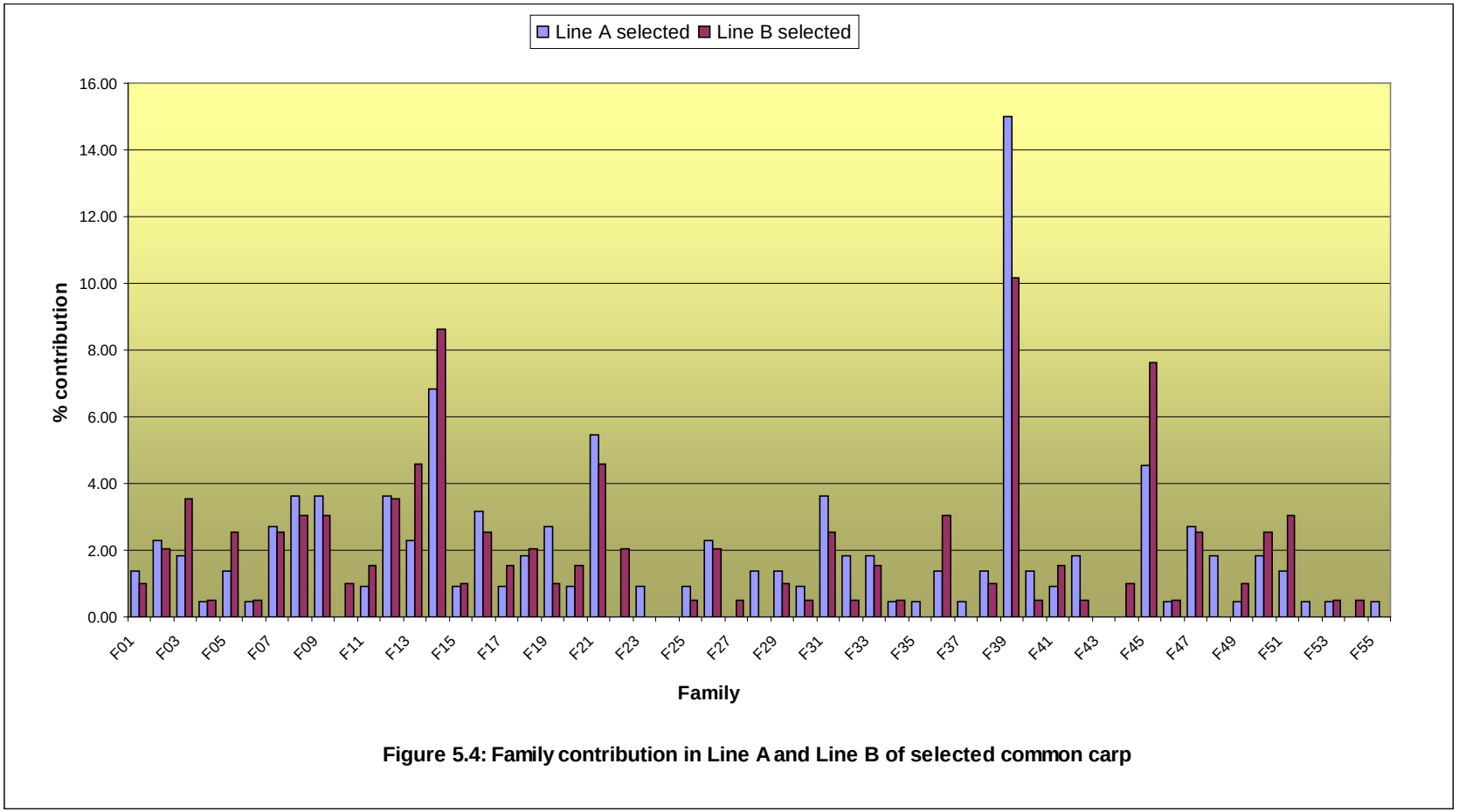

Figure 5.4: Family contribution in Line A and Line B of selected common carp

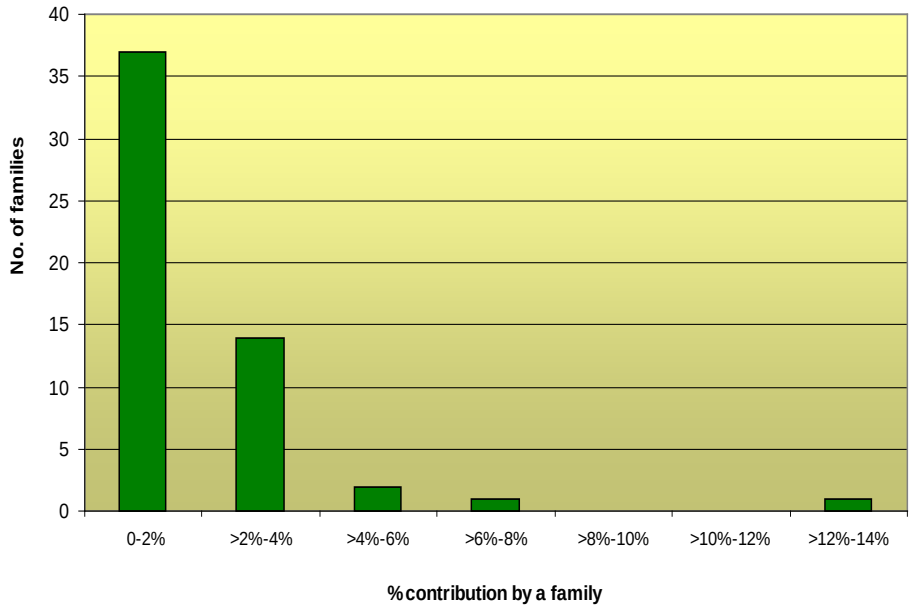

Figure 5.5: Number of families showing different percentage of contribution

Variable family size has had a severe impact on the effective size of the population (N_e), reducing the N_e to less than half of the census size (N) as is observed from the N_e/N values when estimated for separate lines or the entire selected population (Table 5.7). The N_e/N ratios observed were 0.42 in F_1S Line A, 0.49 in F_1S in Line B and 0.45 in the entire F_1S.

Since in the F_1S Line A, 48 families were represented, this was the maximum number of families that could appear among the Line A breeders. In reality however, only 43 families appeared. The size of the families ranged from 1 to 11 offspring per family. The census size of the population was 96 while the N_e was 57.21. It is worth noting that N_e in the Line A breeder group has increased considerably compared to the entire population of Line A even though its census size was smaller than the whole population. This can be attributed to less heterogeneous family representation among the Line A breeders.

5.3.4 Tracking the loss of genetic variation at microsatellite loci

Loss of genetic variation at three stages of the selection programme was studied and compared using changes in allelic and genotypic features at four microsatellite loci. Features that were used as indices of genetic variation were total number of alleles, effective number of alleles, heterozygosity (observed and expected) and allelic diversity. The stages compared were: (i) base population breeders (F_0B), (ii) 1st generation selected population (F_1S Line A and Line B separately and combined) and (iii) breeder group from Line A selected population (F_1SB). The results for each individual locus in these stages are presented in Table 5.8 and overall genetic change across all the loci is summarized in Table 5.9. A general trend of loss of genetic variation was observed at all loci after one generation of selection. Loss of alleles was observed at all loci except MFW12. In F_0B there was a total of 73 alleles at four loci and in the selected population (combining Line A and Line B) there were 70 alleles (corresponding loss of allele is 4.1%) (Table 5.9). Line A however, retained only 67 alleles (loss of allele 8.2%) and Line B retained 68 alleles (loss of alleles 6.85%). The loss of alleles was most prominent in the Line A breeder group (F_1SB). Nine out of 73 alleles (12.32%) were lost at this stage. This obviously has occurred as a result of random drift due to a smaller size of the breeder group. Figures 5.6a-5.6d compares the frequency distribution of alleles at individual loci at different stages. The figures show that the alleles that were lost were present at very low frequencies in the F_0B population. Allelic diversity, A_d, has been used to describe the proportion of alleles retained in a population compared to the base population. Allelic diversity values were lower than 1 in all loci except MFW12.

Table 5.8: Genetic variability at four microsatellite loci in different stages of common carp selection programme

Locus	F_0B	F_1S			F_1SB
		Line A	Line B	Combined	
MFW12					
n	110	234	201	435	126
A	15	15	15	15	15
A_e	9.75	9.14	9.26	9.29	9.7
A_d	-	1	1	1	1
H_o	0.845	0.833	0.831	0.832	0.841
H_e	0.901	0.892	0.894	0.893	0.9
P	>0.001	>0.001	>0.001	>0.001	>0.001
MFW16					
n	107	232	194	426	125
A	20	19	17	19	17
A_e	10.1	9.35	10.75	10.16	8.98
A_d	-	0.95	0.84	0.95	0.84
H_o	0.729	0.677	0.639	0.66	0.672
H_e	0.905	0.895	0.91	0.903	0.892
P	>0.001	>0.001	>0.001	>0.001	>0.001
MFW18					
n	110	234	200	434	126
A	18	15	17	17	15
A_e	7.66	7.21	7.58	7.43	6.52
A_d	-	0.82	0.94	0.94	0.82
H_o	0.755	0.679	0.655	0.668	0.635
H_e	0.873	0.863	0.87	0.867	0.85
P	>0.001	>0.001	>0.001	>0.001	>0.001
MFW29					
n	110	232	201	433	125
A	20	18	19	19	17
A_e	11.9	9.19	11.11	10.54	9.18
A_d	-	0.89	0.95	0.95	0.84
H_o	0.727	0.905	0.871	0.889	0.928
H_e	0.92	0.901	0.913	0.906	0.895
P	>0.001	>0.001	>0.001	>0.001	>0.001

F_0B = base population breeders
F_1S = F_1 selected population
F_1SB = breeders from F_1S Line A
n = sample size
A= number of alleles
A_e = effective number of alleles
A_d = allelic diversity measured against 1st stage (F_0B)
H_o = observed heterozygosity
H_e = expected heterozygosity
P is the probability of Hardy-Weinberg equilibrium test

Table 5.9: Overall genetic changes across loci at different stages of common carp selection programme

All loci together	F_0B	F_1S			F_1SB
		Line A	Line B	Combined	
Total no. of allele	73	67	68	70	64
Mean A	18.20	16.75	17.00	17.50	16.00
Mean A_e	9.85	8.72	9.68	9.36	8.60
Overall A_d	-	0.92	0.93	0.96	0.88
Loss of alleles	-	8.22%	6.85%	4.1%	12.32%
Mean H_e	0.904	0.888	0.897	0.892	0.884
Loss in H_e	-	1.6%	0.7%	1.2%	2%
Mean H_o	0.764	0.774	0.749	0.762	0.769

F_0B = base population breeders
F_1S = F_1 selected population
F_1SB = breeders from F_1S Line A
A= number of alleles
A_e = effective number of alleles
A_d = allelic diversity measured against 1st stage (F_0B)
H_e = expected heterozygosity
H_o = observed heterozygosity

The effective number of alleles (A_e) has also reduced in most cases compared to F_0B. In the F_1S Line B population, however, the A_e at MFW16 (10.75) was greater compared to that in F_0B (10.1). Since values of A_e depend on the homogeneity of allele frequency distribution, this probably has happened as a result of slight homogenization of allele frequency through loss of three rare alleles (Figure 5.9b).

Expected heterozygosity in individual loci as well as across the loci has decreased in selected populations compared to the base population. Mean H_e was higher in Line B (0.897) compared to Line A (0.888), indicating a greater loss of genetic variation in Line A. F_1SB group showed the lowest H_e (0.884) (Table 5.9). The loss in heterozygosity, however, was less prominent compared to the loss of alleles in all stages. The correlation between percent loss of alleles and percent loss of heterozygosity was not found to be significant ($r^2 = 0.543$, $p = 0.26$). However, this should be taken with caution as the sample size was very low (only 4 pairs).

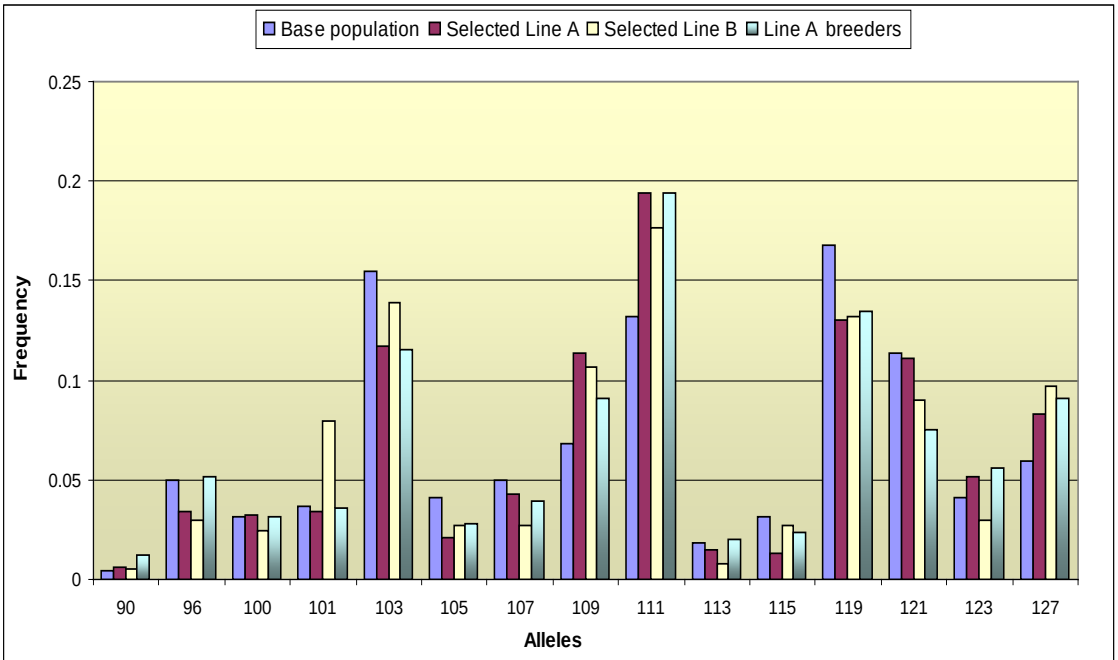

Figure 5.6a: Comparison of allele frequency distribution at MFW12 among base population, two lines of selected population and in breeders from Line A selected population

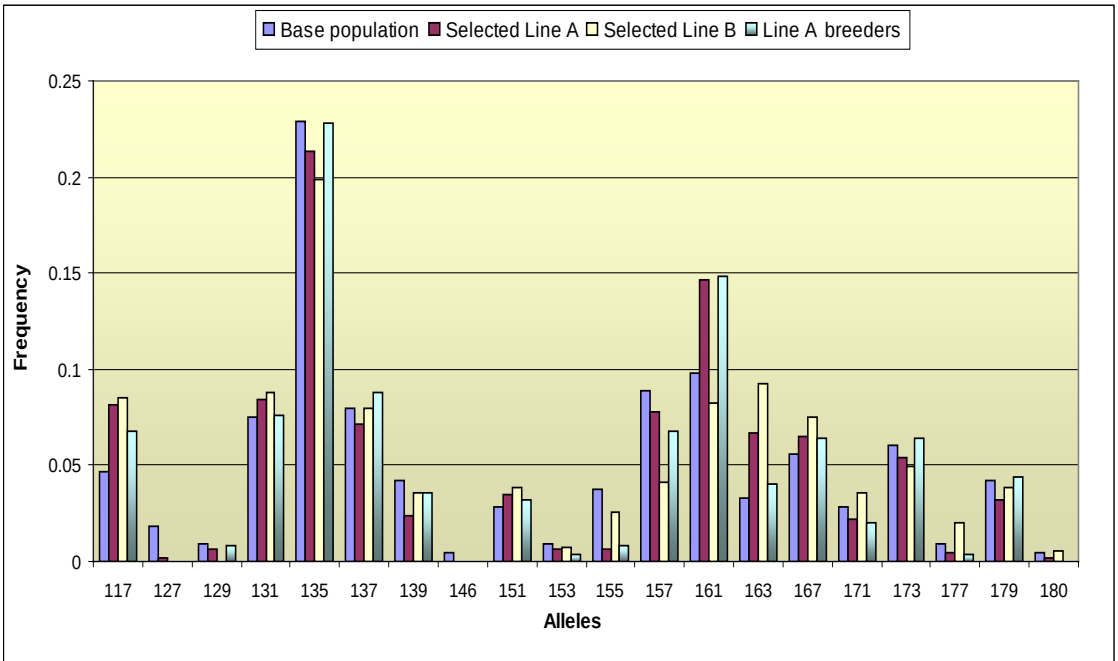

Figure 5.6b: Comparison of allele frequency distribution at MFW16 among base population, two lines of selected population and in breeders from Line A selected population

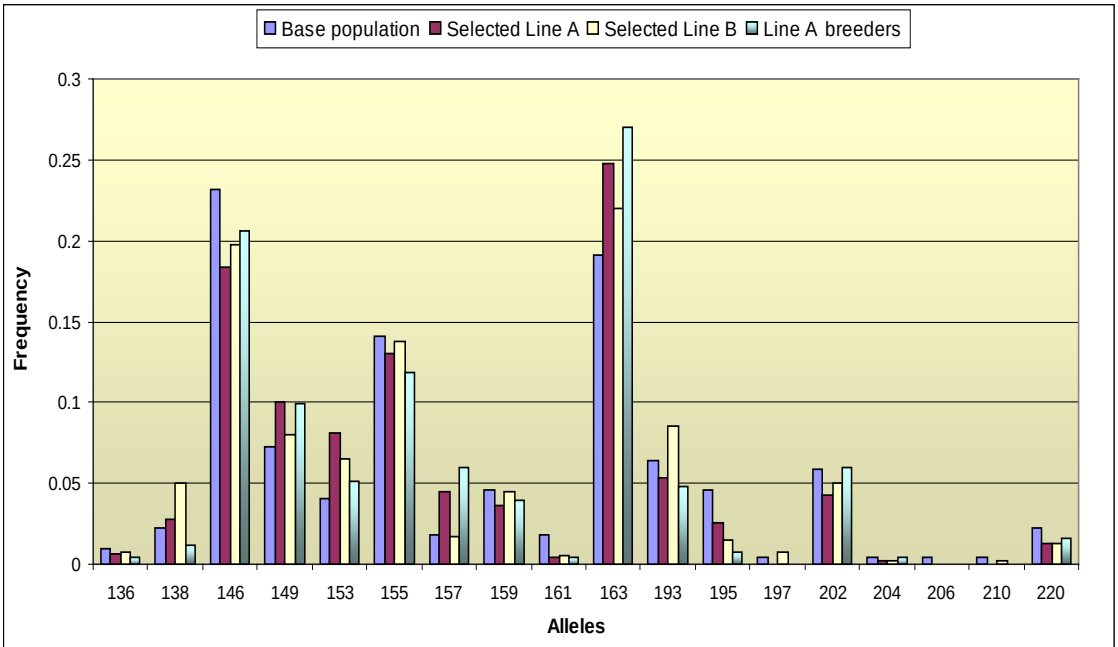

Figure 5.6c: Comparison of allele frequency distribution at MFW18 among base population, two lines of selected population and in breeders from Line A selected population

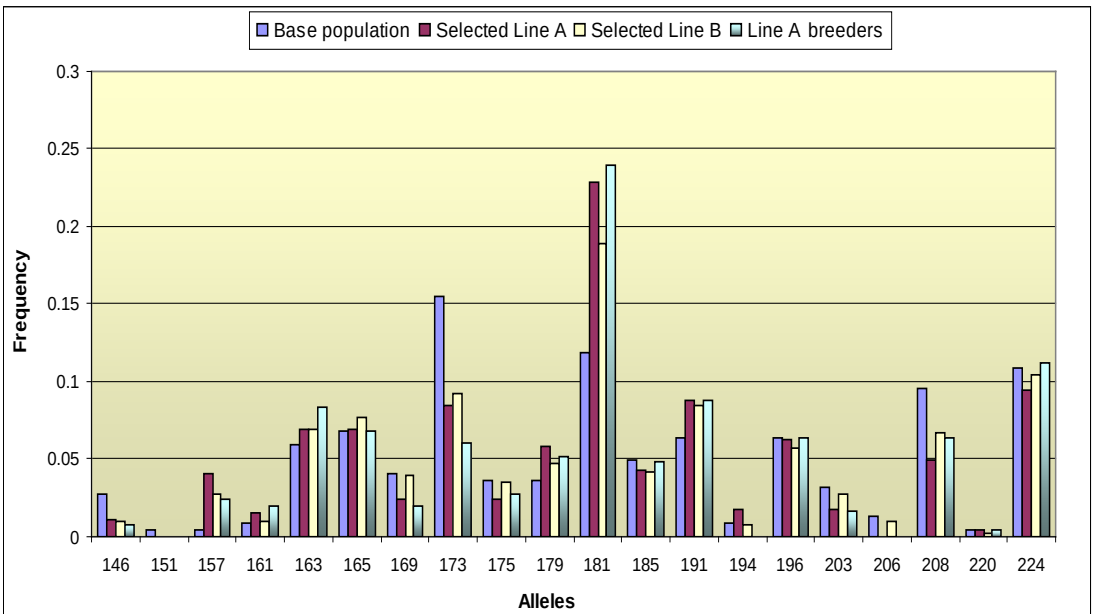

Figure 5.6d : Comparison of allele frequency distribution at MFW29 among base population, two lines of selected population and in breeders from Line A selected population

Observed heterozygosity (H_o) did not correspond with the changes in other features. Therefore, it can be concluded that H_o is not a good indicator of loss in genetic diversity in populations; rather it can reflect some other features such as distribution of genotypes.

The change in genetic structure among different groups was studied by comparing their allelic and genotypic distributions. The results of comparison of allelic and genotypic distributions among the groups or stages are presented in Table 5.10 and Table 5.11 respectively. Allelic distributions in all three stages of the selected population (F_1S Line A, F_1S Line B and F_1SB) were found to be significantly different compared to the allelic distribution in the base population breeders (F_0B). Line A breeders (F_1SB), however, did not show significant difference in allelic distribution from the entire Line A selected population or from Line B selected population. For genotypic distribution only the Line A selected fish showed significant deviation (p= 0.00068) from base population breeders. Even though the Line A breeders were chosen randomly from the Line A selected population, its overall genotypic distribution was not significantly different from the base population breeders (F_0B).

Table 5.10: Result of pairwise comparison of different stages of common carp selection programme for difference in allelic distribution across four loci.

Population pair compared		χ^2 value	df	P-value
Base population breeders (F_0B)	F_1 selected Line A (F_1S Line A)	Infinity*	8	Highly sign.
Base population breeders (F_0B)	F_1 selected Line B (F_1S Line B)	16.647	8	0.034
Base population breeders (F_0B)	F_1 selected breeders Line A (F_1SB Line A)	18.61	8	0.01709
F_1 selected Line A (F_1S line A)	F_1 selected Line B (F_1S Line B)	17.299	8	0.02714
F_1 selected Line A (F_1S line A)	F_1 selected breeders Line A (F_1SB Line A)	0.862	8	0.99898
F_1 selected Line B (F_1S line B)	F_1 selected breeders Line A (F_1SB Line A)	14.248	8	0.07552

***** The result was given as such by the programme GENPOP

Table 5.11: Result of pairwise comparison of different stages of common carp selection programme for difference in genotypic distribution across four loci.

Population pair compared		χ^2 value	df	P-value
Base population breeders (F_0B)	F_1 selected Line A (F_1S Line A)	27.101	8	0.00068
Base population breeders (F_0B)	F_1 selected Line B (F_1S Line B)	11.422	8	0.17892
Base population breeders (F_0B)	F_1 selected breeders Line A (F_1SB Line A)	14.495	8	0.06974
F_1 selected Line A (F_1S line A)	F_1 selected Line B (F_1S Line B)	11.525	8	0.17368
F_1 selected Line A (F_1S line A)	F_1 selected breeders Line A (F_1SB Line A)	0.857	8	0.999
F_1 selected Line B (F_1S line B)	F_1 selected breeders Line A (F_1SB Line A)	10.116	8	0.25696

5.4 Discussion

Maintaining genetic diversity at an acceptable level is crucially important in a selective breeding programme. Loss of genetic variance in populations under artificial selection results from two major causes. In the first generation, the reduction in variance is the result of gametic phase disequilibrium (called the Bulmer effect) (Falconer and MacKay 1996). This happens because the group of selected parents represents one tail of the phenotypic distribution and in consequence their phenotypic variance is less than that of the whole population from which they are selected. Mating between these selected parents causes non-random breeding and gametic phase disequilibrium. In subsequent generations the effect of gametic phase disequilibrium reduces but the loss of variance results from accumulation of inbreeding (Falconer and MacKay 1996; Gjerde *et al.* 1996). Inbreeding tends to counteract the gain achieved though selection process by inbreeding depression, where increased homozygosity among individuals elevates the chance of expression of lethal or lower fitness recessive genes (Falconer and Mackay 1996). The problem of inbreeding is even more pronounced when the selection is being performed on highly fecund animals like fish and shellfish because their high fecundity permits high selection intensities (Gjerde *et al*. 1996; Pante *et al.* 2001), which tends to mean that only a small number of breeding individuals are sufficient to contribute all of the progeny in succeeding generations, leading to increased inbreeding. Genetic management of a selection programme thus requires proper designs for breeding, rearing and selection along with monitoring of the changes in different parameters of genetic variability. The present study involved application of microsatellite markers to monitor the changes in various early stages in a common carp selection programme which is currently under progress in Karnataka, India.

5.4.1 Heterogeneous base population

Sufficient genetic variability is a pre-requisite for success in selective breeding. It is therefore, suggested to start a selection programme from a base population containing a large amount of genetic variability (Allendorf and Ryman 1987; Ferguson 1995). Securing wide genetic variation in the base population is important on several accounts. It allows for the selection of the most desirable phenotype. A high level of genetic variation (combined with breeding practices that minimize the loss of variation during selection) ensures gain for several generations of selection, before any loss of genetic diversity and inbreeding show negative impacts. Finally, the broad genetic profile of the base population keeps open the window for selection of additional traits in future in the improved population (Bentsen and Olesen 2002). A review by Vandeputte (2003) shows that selective breeding for fast growth in common carp had not been proven very effective in most of the previous well-documented attempts. Among many potential causes for the lack of success in common carp selection, low genetic variability of the domesticated stocks was thought to be a major factor.

Several population genetics studies involving common carp strains from different parts of its geographical range have established that there exist two subspecies of common carp, *Cyprinus carpio carpio* from Europe and *C. carpio haematopterus* from Asia (Vandeputte 2003; Kohlmann *et al.* 2003). Vandeputte (2003) suggested that founder stocks for selection programmes with high genetic variability can be synthesized by combining the two subspecies.

The common carp selection programme that is the focal point of present study was started from a base population synthesised from six stocks of different origins and histories. Four of these stocks, Amur, P3, RJ and SV came from exogenous sources

(Hungary, Indonesia and Vietnam) whereas the L-BRP and L-FRS were "local" stocks (these stocks also originated from exogenous stocks) of Karnataka, India. The six stocks contributed different proportions in the base population, which was decided based on their performance in previous growth trials. The high variability of the base population was confirmed by the use of five microsatellite markers. The number of alleles per locus ranged from 15-20 with a mean expected heterozygosity of approximately 90%. Random crossings among the stocks were carried out to create progeny of the first generation from where the first round of selection was performed. Inter-stock crossing incorporated the gene pool from all the stocks, reducing the risk of total loss of the gene pool of smaller contributing stocks.

Microsatellite based parentage analysis was used to monitor changes in representation of stocks at different stages of the selection programme. The current analysis showed that from the very beginning of the selection programme the original representation of the stocks altered. The representation of one major contributor stock, Amur, dropped alarmingly in F_0B (from 25% in the entire base population to 10.91%) as many of the breeders (45%) randomly chosen from the base population failed to breed, probably because these fishes were not mature when the breeding was performed. The percent representation of this specific stock further reduced in the selected group of fish (10.91% in F_0B to 5.28%).

Several potential causes might have been responsible for the large reduction of Amur contribution in the selected population. The potential causes are higher relative mortality of the Amur progeny in the F_1 population, lower growth performance of the progeny so that they tend not to be selected or failure to assign parentage of Amur progeny present in the selected population. Fertilization and hatching success cannot be

considered as causative factors because equal number of fry from each cross was taken and reared. It is, difficult to certainly verify the presence of or the relative magnitude of suspected causes. During the rearing of the F_1 progeny prior to selection, there was large mortality in some of the ponds. As a result of this it was not even possible to select as many fish as was originally intended (the plan was to select 300 fish from each line but only 277 fish in Line A and 201 fish in Line B could be selected). Therefore, this could be a major reason for low representation of the Amur stock in the selected population.

The hypothesis of lower growth performance of the Amur progeny is difficult to assess. The Amur stock showed the best growth performance in the stock evaluation trials that were conducted before the commencement of the selection programme (as explained in section 5.1.1). However, it is important to note here that due to random choice of breeders to form the mating pairs, many of the crossings were between individuals from two different stocks, that is, the progeny were hybrids rather than pure bred. In the evaluation trials it was observed that three inter-stock hybrids suffered from slightly negative heterosis in growth rate and two of them were Amur x P3 cross (relative heterosis was -9.71%) and Amur x L-BRP cross (relative heterosis -2.98%) (unpublished data from Basavaraju *et al*.). To investigate if this was a cause of poor growth performance of Amur-descended progeny, individual crosses were checked. There were a total of 7 crosses where one or both breeders were from Amur stock. Four of these were pure cross (Amur x Amur) and the other three were inter-stock crosses. Among the inter-stock crosses, one was P3 x Amur cross and one Amur x L-BRP cross. Table 5.12 shows the number of offspring produced by different Amur crosses. All the pure Amur crosses had very few offspring (0-5). Since the evaluation trials showed that pure Amur offspring had the greatest growth rate, therefore, for these pure crosses

greater mortality of offspring is perhaps a more plausible explanation rather than poor growth.

Table 5.12: Representation of Amur-descended families in the selected population

Family	Cross	No. of offspring*
Fam 03	Amur X L-FRS	11
Fam 16	Amur X SV	12
Fam 23	P3 X Amur	2
Fam 42	Amur X Amur	5
Fam 43	Amur X Amur	0
Fam 49	Amur X L-BRP	3
Fam 52	Amur X Amur	1
Fam 53	Amur X Amur	2

* Expected no. of offspring/family was 7.95

However, there may be others factors such as environmental stress and competition with other individuals which might have reduced the growth of pure-bred Amur offspring. The P3 x Amur cross and the Amur x L-BRP cross also showed very few offspring (2 and 3 respectively) for which negative heterosis can be one of the reasons. However, during the evaluation trial P3 x L-BRP cross also showed negative heterosis but it did not affect crosses involving these stocks (result not shown here). Therefore, the negative heterosis cannot be given much weight as a major factor.

Finally, the percent contribution of stocks in the selected population was calculated only on those fish which could be assigned to a single family. In the present study, the parentage assignment success was slightly above 95%. It might be if the other 5% could be assigned, the representation would have increased for Amur stock. However, even if all the non-assigned offspring were the hybrid progeny of Amur with some other stock, the maximum representation would be around 7%, still much lower than the expectation (11%).

5.4.2 Family contribution, effective population size and inbreeding

Mass selection programmes are particularly notorious for allowing rapid accumulation of inbreeding. In these breeding programmes selection is carried out based on individual's performance. Generally no pedigree record is used to avoid breeding between close relatives or in recruiting breeders for the successive generations. This allows superior families to contribute more offspring to the next generation than the average families (Bijma *et al*. 2000). Differential representation of families and unequal sex ratios are two of the major causes of reduction of effective population size (N_e) and thereby increase the rate of inbreeding. According to Vandeputte (2003) one of the possible reasons of low response in some common carp selection programmes, such as the one reported by Moav and Wohlfarth (1976), was the differential contribution of the breeders.

The selective breeding programme in the present study used pairwise mating as a means to mitigate the problem of inbreeding due to unequal sex ratio and unequal family size. An equal amount of fry was taken from every breeding pair. The use of microsatellite markers to assess the family contribution showed that out of 55 potential full-sib families, 53 families were represented in the selected population, which is a remarkably high number. This highlights the usefulness of pairwise mating in mass selection. The detrimental effect of mass spawning on N_e as against pairwise mating has been reported in many species such as gilthead seabream (Brown *et al.* 2005), Japanese flounder (Sekino *et al.* 2003) and has been discussed in detail in previous chapter (Chapter 4). Even though most of the families contributed to the selected population, the family sizes were highly skewed, ranging from 1 to 53 offspring. Due to this highly unequal family representation the N_e of the selected population dropped to 49.03 which is less than half

of the census population size (110). An inbreeding rate (ΔF) of 1.09% is expected at this level of N_e as was calculated from $1/2N_e$.

What should be the acceptable maximum rate of inbreeding and minimum acceptable N_e are matters of debate. Gjerde *et al.* (1996) suggested that the tolerable rates of inbreeding may differ among populations and among traits depending on the inbreeding depression in production traits and fitness. Thus acceptable rates of inbreeding may vary from one breeding programme to another. Meuwissen and Woolliams (1994) showed that in the case of livestock populations under selection the minimum N_e required to avoid a decline in fitness after ten generations of selection ranges from 31 (ΔF=1.6%) to 250 (ΔF=0.2%) per generation, depending on the heritability of the trait. Bentsen and Olesen (2002) have shown through stochastic simulation carried out on a 15 generation of mass selection programme, that to keep the inbreeding rates low (about 1% per generation) a minimum of 50 pairs of breeders (i.e. N_e=100) should be selected. This number is double the number of breeders required to keep the same level of inbreeding in a random mating population without selection as shown by Falconer and Mackay (1996). The N_e level (49.03) and inbreeding rate (1.09%) observed in the present experiment taking into account variable family size, is in the acceptable range suggested by Meuwissen and Woolliams (1994). However, this N_e is much smaller (less than half) than that suggested by Bentsen and Olesen (2002). This is probably due to the fact that Bentsen and Olesen (2002) carried out a stochastic simulation where the effect of differential family contribution was not addressed. If the family size was not a concern in the present study the N_e would have been equal to the number of total breeders and hence would be 110 (as 55 pairs eventually bred).

The drastic drop of N_e due to unequal contribution of families can be rectified by choosing equal number of individuals from each family as breeders for the next generation as far as possible. This would however, require precise pedigree information, which is available in the present case due to the use of microsatellite markers and PIT tags for pedigree tracing.

5.4.3 Assessment of loss of genetic diversity using microsatellites

Molecular markers viz. allozymes, mtDNA and microsatellites have been extensively used in assessing the genetic diversity of populations. In recent years the application of microsatellites has surpassed the use other markers for genetic diversity studies owing to their abundance and high levels of variation. In many studies microsatellites have been effectively used in detecting loss in genetic variation in captive populations compared to wild populations (Norris *et al.* 1999; Xu *et al.* 2001; Was and Wenne 2002; Koljonen *et al.* 2002; Kohlmann *et al.* 2003; Alarcón *et al.* 2004; Evans *et al.* 2004). While the use of microsatellites has increased, so has the controversy about the acceptance of microsatellite-assessed genetic diversity as a measure of fitness of a population. This is because of the fact that microsatellites are selectively neutral markers. It is thus thought that they may lose genetic variation more rapidly than the loci concerned with fitness (Reed and Frankham 2003). Reed and Frankham (2001) performed a meta-analysis to investigate the suitability of molecular heterozygosity (measured by molecular markers) as a surrogate for the measures of quantitative genetic variations such as heritability, coefficient of variation of phenotypic traits and genetic distance. They observed moderate and significant correlation between heterozygosity and the co-efficient of phenotypic variation ($r = 0.36 \pm 0.07$), a weak but significant correlation between heterozygosity and genetic distance ($r = 0.22 \pm 0.08$) but non-significant correlation between heterozygosity and heritability ($r = -0.08 \pm 0.11$). The

mean correlation between molecular and quantitative measures of genetic variation was weak (r = 0.217) and they concluded that molecular measures of genetic diversity have only a very limited ability to predict quantitative genetic variability. In a more recent meta-analysis, Reed and Frankham (2003) showed that there exists a moderate correlation between population fitness (measured as the number of adult progeny produced or population growth rate) and molecular heterozygosity (r = 0.447 ± 0.081).

In spite of all the controversies, molecular markers can still offer useful information about changes in genetic diversity, even though this may not directly reflect changes in population fitness or quantitative genetic variations. Microsatellite markers were used on the selective breeding programme of common carp to monitor the changes in genetic diversity during the selection process. Total number of alleles (A), effective number of alleles (A_e), allelic diversity (A_d), and heterozygosity values (H_e) were studied as indices of genetic diversity. All these features showed loss in genetic diversity in the 1st generation of the selected population. As a consequence of selection, 4.1% of alleles were lost in the 1st generation of selection. In the separate replicate lines of selected fish the loss was even higher (6.85% to 8.21%) due to smaller size of the populations. Alleles that were lost were mainly present in low frequency. Thus random drift was the more likely driving force for the loss of alleles than the selection process itself. In the Line A breeders the number of alleles observed was reduced further showing a 12.32% loss. Loss of heterozygosity was less pronounced than the loss of alleles. Many other studies using microsatellites for genetic differentiation suggested that loss of alleles or allelic diversity is more effective measure for reduction in genetic variation than heterozygosity values (Coughlan *et al.* 1998; Norris *et al.* 1999; Perez-Enriquez *et al.* 1999; Was and Wenne 2002). For instance, a study on several hatchery reared abalone populations showed a large decline in the number of alleles (35-62%) across several

microsatellite loci but did not show any associated loss of heterozygosity (Evans *et al.* 2004).

All the selected groups (Line A, Line B and Line A breeders) showed significant variation in allelic distribution compared to the base population. For genotypic distribution, Line A showed highly significant variation from the parental population but Line B did not show statistically significant deviation. The significant differences in allelic and genotypic distributions in the first generation selected lines indicate that they are becoming genetically distinct from their origin. Moreover, since Line A and Line B will be maintained separately and selection will be carried out within the lines, this will ultimately lead to creation of distinct strains. In the present study the two lines were found to be significantly different in allelic distribution but not in genotypic distribution.

5.4.4 Recommendations

The present study shows that microsatellites can be a very useful tool in monitoring different genetic aspects in a selective breeding programme. Such monitoring can help in planning suitable methods to reduce rate of loss of genetic diversity. In the present study, the drastic loss of Amur stock from the selected population and rapid reduction of N_e due to unequal family size were detected as major problems. To improve these situations the following two recommendations can be forwarded:

1. To improve the Amur representation, some Amur broodfish from the base population (taking the biggest individuals as possible) could be crossed with the broodfish from F_1 selected population. If the loss of Amur in previous case was not due to low fitness, the reintroduction of Amur in the F_2 generation might solve the problem.

2. To overcome (or reduce) the problem of highly variable family size and rapid reduction of N_e, same (or similar) number of fishes from each family could be selected for breeding. The principle of combined selection - the best individuals from best families - can be applied here.

CHAPTER 6

ISOLATION AND CHARACTERIZATION OF MICROSATELLITE MARKERS FROM A ROHU (*Labeo rohita*) GENOMIC LIBRARY

6.1 Introduction

Labeo rohita, commonly called rohu, is a cyprinid species native to South Asia and is naturally distributed in the rivers of Pakistan, northern India, Bangladesh, Myanmar and Nepal (FishBase 2005). It is also one of the most popular aquaculture species in this region. Among the carps cultured in Asia, rohu production was ranked 7th (713267 tonnes) in 2003 (FAO 2005). In Bangladesh it contributed about 27% of the total aquaculture production (FAO 2005). Rapid expansion of rohu aquaculture has been possible because of the availability of the hatchery produced seed. Currently about 99% of the rohu seed originates from private and public hatcheries (Islam and Alam 2004). Rapid intensification of fry production in hatcheries however, has not been accompanied by any genetic quality maintenance. Apart from the problems like inbreeding and negative selection, indiscriminate hybridization of rohu with catla and mrigal has been identified as a major problem and is assumed to be partly responsible for the decreased growth performance of the species in aquaculture (Simonsen *et al.* 2004). Introduction of hatchery originated poor quality rohu seeds to the wild through escapes and government sponsored seed stocking programme, often on a massive scale, are becoming a serious concern (Islam and Alam, 2004). It has therefore, become imperative to understand the genetic composition of natural rohu populations in order to ensure their proper management. Population genetics studies such as identification of natural stocks and detection of changes in genetic make-up of wild populations as a result of mixing of hatchery fish can only be performed with the aid of molecular markers. Moreover, it has become urgent to use marker technology to identify the inbreeding level of rohu stocks in hatcheries, estimate the relatedness between individuals and stocks and for other on-farm management related issues. The importance of molecular markers, especially microsatellites, in acquiring pedigree

information and its implications in avoiding inbreeding and in communal rearing has already been discussed in previous chapters of the present thesis. So far only a few genetic markers, mainly allozymes (Alam *et al.* 2002) and RAPD (Islam and Alam 2004) have been used for rohu population related studies. Very recently, 12 microsatellite markers were reported by Das *et al*. (2005). The present study reports new microsatellites isolated from a partial genomic library of rohu.

6.1.1 Available methods of isolation of microsatellites

Microsatellites for a species can be isolated using four approaches: (i) by searching public databases on DNA sequences; (ii) by cross-species amplification of primers from related species; (iii) by screening genomic libraries and (iv) by using non-library protocols by screening RAPD profiles (Maguire 2001). Database searching is the least expensive means and hence can be used as a starting point for isolating microsatellite markers, provided there are sufficient entries in the database. For instance, this technique has been used in tiger puffer *Takifugu rubripes* and 12 microsatellite loci have been isolated (Takagi *et al.* 2003). Database searching, however, alone is unlikely to generate sufficient markers especially when they are needed in large numbers such as for genetic mapping. Besides a large proportion of the sequences available in the database are "Expressed Sequence Tags" (EST). Although microsatellites can be generated from the EST sequences, the problem is that they are generally trinucleotide repeats and are less polymorphic than dinucleotide based microsatellites (Cho *et al.* 2000). Hence these have lower values for population studies than those isolated from genomic libraries.

Cross-species amplification has been used to develop microsatellites in many species of fish. The success of deriving microsatellites through this method, however, largely

depends on the taxonomic distance between the species of interest. Conservation of the flanking sequence determines whether the correct region is amplified, and how variable the microsatellite loci are. Often, the PCR reaction conditions need to be optimized and the products sequenced to verify the presence of a microsatellite region. Moreover, many heterologous primer pairs may need to be screened to find useful primers for the species of interest. This method may be cost-effective if there is access to many primer pairs from related species. Recently, microsatellites for a number of cyprinid species in South Asia have been developed using cross species amplification approach, e.g. in mrigal (Lal *et al.* 2004), *Tor putitora* (Mohindra *et al.* 2004) and *Labeo dero* (Mohindra *et al.* 2005).

Traditionally microsatellite loci have been isolated from partial genomic libraries, screening several thousands of clones through colony hybridization with repeat containing probes. Although the procedure is relatively simple for microsatellite-rich genomes, this approach can turn out to be extremely tedious and inefficient for species with low microsatellite frequencies, such as birds or plants, or when large number of markers are required (Zane *et al.* 2002). The number of positive clones (containing microsatellite repeats) that can be obtained by this traditional method usually ranges from 12% to less than 0.4%. Therefore, several alternative strategies have been devised in order to reduce the time invested in microsatellite isolation and to significantly increase the yield. These include different enrichment protocols for increasing the frequency of the repeat containing inserts in the genomic library to be screened for microsatellites (Karagyozov *et al* 1993; Armour *et al.* 1994; Kijas *et al.* 1994; Ostrander *et al.* 1992; Paetkau 1999).

To avoid the time consuming tasks of library construction and screening, some non-library PCR based methods have been proposed. These are based mainly on a modified RAPD approach where the microsatellites are screened either by using repeat-anchored random primers or by using conventional RAPD primers followed by Southern hybridization of PCR bands with microsatellite probes (Zane *et al.* 2002 and references therein). While all these methods provide, if successful, a quick alternative to laborious and time consuming library screening, their use has not been very frequent.

Since for isolation of rohu microsatellites in the present work the library screening approach was adopted, different core protocols for genomic library screening are briefly described in the sections below, after the review papers of Maguire (2001) and Zane *et al.* (2002).

6.1.1.1 Traditional approach of library screening

In the traditional protocol, a partial genomic library of the target species is created by fragmenting high quality genomic DNA, usually by restriction enzymes or less frequently by sonication. The length of DNA fragments produced by sonication is less dependent on genomic nucleotide composition. On the contrary, when restriction enzymes are used the average length of DNA fragments depends on base composition of the genome and the recognition sequence of the enzymes used. Differences in nucleotide composition within genomes, therefore, may lead to unequal sampling of genomic regions. To overcome this limitation, digestions by several restriction enzymes have been proposed. Multiple digestions can be carried out either simultaneously, which results in a smaller average size of the fragments, or performed separately and then pooled together thereby producing longer fragments.

After digestion of the genomic DNA, fragments of the desired sizes are isolated by running the DNA on an agarose gel. Generally fragments in the range of 200-1000 bp are ideal for successful cloning and in recovering enough flanking regions to design primers for the amplification of individual microsatellites. Depending on the type of endonuclease used, the digested fragments have either sticky or blunt ends. The successive ligation step depends on the nature of the termini of fragmented DNA. Many different processes have been described for ligation, ranging from cohesive-end ligation into a dephosphorylated plasmid (Kijas *et al.* 1994) to linker mediated ligation of blunt-ended DNA into a dephosphorylated λgt10 vector (Kandpal *et al.* 1994). Upon ligation of the DNA fragment with a suitable vector, competent bacterial cells are transformed with the ligation product. Culture of the transformed bacterial cells yields thousands of recombinant clones that are then screened for the presence of microsatellite sequences. Screening of positive clones is carried out by means of Southern hybridisation using repeat-containing probes, after blotting the bacterial colonies onto nylon membranes. Transfer of colonies to membranes can be carried out either by replica plating or by picking single colonies and ordering them in new arrayed plates. Although the latter method is more time consuming and hence limits the number of clones that can be screened, it avoids the requirement of reprobing positive clones for confirmation. Probes can be labelled by both radioactive (^{32}P, ^{33}P) and non-radioactive (digoxigenin) methods. Radioactive protocols are more sensitive but require stringent lab conditions for handling radioisotopes. Furthermore, the short life of radioisotopes poses limitations on their use. Non-radioactive labelling techniques allow less stringent and safer working conditions and also allow long-term storage of the probes.

Following sequencing of repeat-containing clones, specific primers are designed and PCR conditions optimized to allow the amplification of each locus from different individuals of a population.

6.1.1.2 Enrichment protocols

For isolating a large number of microsatellite repeats different types of 'library enrichment' protocols have been developed. Two major protocols are: selective hybridisation (Karagyozov *et al.* 1993; Armour *et al.* 1994; Kijas *et al.* 1994) and primer extension prior to cloning (Ostrander *et al.* 1992; Paetkau 1999).

Selective hybridization approach

Selective hybridization protocols are extremely popular for their straightforward strategy. The application of this approach increased the rate of positive colonies (from 20% to 90%) in a large variety of taxa, from plants to vertebrates, using di, tri-, and tetranucleotide probes (Zane *et al.* 2002). Like the traditional approach the "selective hybridisation" also begins with creating small genomic fragments and then ligating them with vectors or adaptors. PCR amplification is then carried out on the ligated products to obtain a sufficient amount of DNA for selective hybridization. Selective hybridization is performed using specific repeat-containing oligonucleotides serving as probes. The probe can be bound to a nylon membrane or can be biotinylated at the 5′ end and bound to streptavidin coated beads. The use of biotinylated probe is generally preferred because in the liquid medium the probe is fully available for hybridization. In contrast the nylon bound probe is partially cross-linked to the membrane, and therefore hybridizes less efficiently with the target DNA. Upon hybridization several washes are performed to remove non-specific binding. The DNA is then eluted and recovered by PCR amplification. Finally, the enriched DNA is cloned into a suitable vector to create

a repeat-enriched library. Depending on the efficiency of the whole procedure recombinant clones can be directly sequenced or screened for the presence of repeats by Southern blotting and probing. Figure 6.1 presents a schematic diagram of selective hybridization.

Primer extension approach

Two protocols for primer extension enrichment of microsatellites repeats have been proposed: one by Ostrander *et al.* (1992) and the other by Paetkau (1999). These methods were reported to be very efficient for the enrichment of dinucleotide repeats (mainly applied for AC repeats) yielding from 40 to 50% (Ostrander *et al.* 1992) up to 100% of positive clones (Paetkau 1999). These protocols involve a rather large number of steps which has limited their use. In both the methods a 'primary' genomic library is constructed by inserting fragmented genomic DNA into a Phagemid or Phage vector in order to obtain a single stranded DNA (ssDNA) library. The ssDNA serves as template for primer extension reactions where repeat-specific oligonucleotides are used as primers to generate double-stranded products only from vectors containing the desired repeat. The two enrichment procedures differ in the strategies used to recover primer-extended products (Fig 6.2). In their protocol, Ostrander *et al.* (1992) eluted 40,000-60,000 colonies from a phagemid library cultured on LB-agar plates, grew them to saturation in liquid media and superinfected with M13 helper phage. Because of the particular genotype of the bacterial host (dut^-ung^- which is deficient in dUTPase and uracil-*N*-glycosylase), superinfection resulted in a library of circular ssDNA containing uracil instead of thymine. Single stranded DNA (ssDNA) was selectively converted to double stranded DNA through $(CA)_n$ or $(GT)_n$ primer extension and the mixture was used to transform a dut^+ung^+ *Escherichia coli* strain. The resulting library was highly enriched for repeat-containing inserts because only the double stranded products were

ultimately retained after transformation. If any bacteria were transformed with native single strand products (which did not have specific repeats) the uracil-N-glycosylase (*ung*$^{+}$) degraded them due to the presence of uracil. In contrast, the double-stranded DNA products were rescued because the thymidine-containing primer-extended strand allowed for the action of host repair mechanisms.

In the Paetkau protocol, the primary library was obtained using M13 phage, and circular ssDNA was obtained through elution of 30,000 clear plaques. Primer extension was performed using 5$^{/}$ biotinylated repeat-specific oligonucleotides and Klenow DNA polymerase. This resulted in a population of circular DNAs whose second strand was a linear molecule with biotin at one end. Biotinylated products were selectively recovered from the reaction mix using streptavidin-coated beads and after washing steps, circular phage ssDNA was released by denaturation. Finally, molecules containing the microsatellites were converted to double stranded molecules with a second round of primer extension and were used for the final transformation.

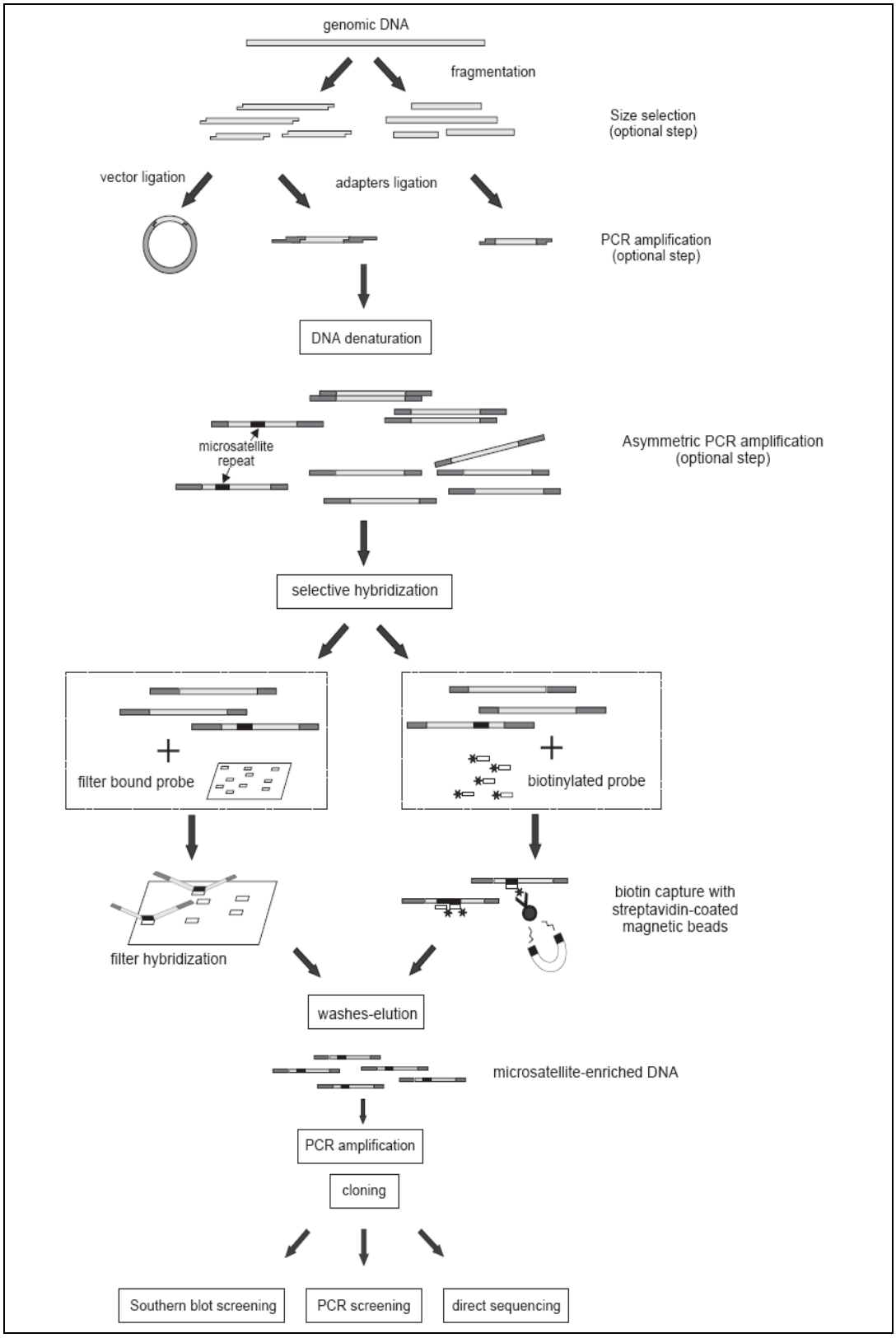

Figure 6.1: Schematic diagram of selective hybridization protocols for enrichment of microsatellite motifs (taken from Zane *et al.* 2002)

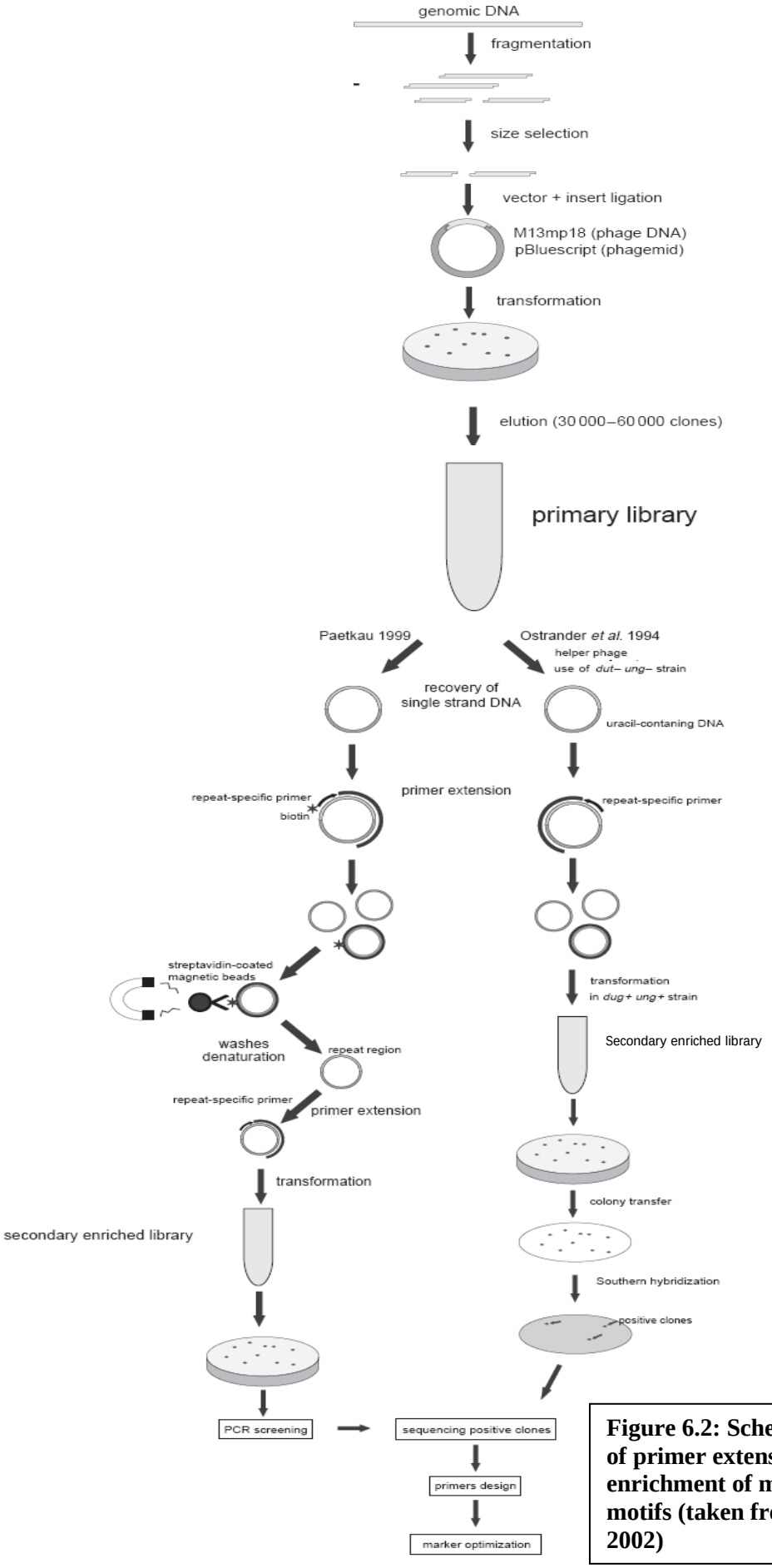

Figure 6.2: Schematic diagram of primer extension protocols for enrichment of microsatellite motifs (taken from Zane *et al.* 2002)

6.1.2 Objectives of the study

The objectives of the present study were:

1. To isolate a number of microsatellite markers from rohu genomic DNA using an enrichment protocol.
2. To check the polymorphism of the isolated microsatellites in wild rohu samples.
3. To characterize the polymorphic microsatellite loci.

Cross-amplification of some of the newly isolated microsatellites in other cyprinid species was also assessed in collaboration with an MSc project student.

6.2 Materials and Methods

Isolation of microsatellites from rohu genome was carried out using selective hybridization protocol of Kijas *et al.* (1994) with slight modifications. The modifications were suggested by Dr. Paulo A. Prodöhl of Queens University, Belfast.

6.2.1 Construction of primary library

Construction of the primary library consisted of four steps: digestion of genomic DNA with restriction endonuclease, selection of desired fragment size range, preparation of vectors for cloning and ligation of the size selected fragments into vectors.

6.2.1.1 Digestion of genomic DNA

Dr. John Taggart of the Institute of Aquaculture, University of Stirling started creating genomic libraries from a number of carp species. As a part of this he digested the rohu genomic DNA which was later used for microsatellite isolation in the present work. Dr. Taggart used high quality DNA extracted from heart tissue of four individuals using a phenol-chloroform protocol. DNA from all four individuals was subsequently pooled. Digestion was carried out in two sets. In one set digestion was performed separately with three 4-base recognition restriction enzymes (*Alu* I, *Hae* III and *Rsa* I) and the digested fragments were later pooled. The second set was created by simultaneous digestion of genomic DNA with five 6-base recognition restriction enzymes (*Dra* I, *Eco*R V, *Hpa* I, *Nae* I and *Pvu* II). Both the sets of digested DNA were suspended in TE buffer and preserved. From here on the two sets of digested products will be referred to as D4 and D6 respectively. The work was taken over from this stage to carry on with the microsatellite isolation.

6.2.1.2 Fragment size fractionation

For size fractionation 40 µl of digested product was run on a 1.5% agarose gel at low voltage (25 volt) for 1hr 15 minutes. Size marker φ174 *Hae* III was run alongside to help identify the desired size range of the fragments. Bands (smear) corresponding to the approximate size between 200-700 bp were excised from the gel. DNA fragments from the gel slice were purified using the GFX^{TM} PCR DNA and Gel Band Purification Kit (Pharmacia). The purified fragments were finally eluted in 2 mM Tris buffer (pH 8.0) and quantified by techniques described in Chapter 2 (section 2.2).

6.2.1.3 Preparation of vector for blunt cloning

The vector pBluescript II KS (-) phagemid (map and characteristics described in Figures 6.3 and 6.4 and Table 6.1) from Stratagene was used to create the primary library with the digested products. The vector was prepared for blunt end cloning by digesting with excess *Eco*R V restriction enzyme followed by dephosphorylation using Calf Intestinal Phosphatase (CIP) (the prepared vector was supplied by Dr. John Taggart and hence a detailed description is not given here).

6.2.1.4 Ligation

The insert-vector amount for the ligation was determined by the following formula:

$$\frac{ng\ of\ \ vector \times kb\ size\ of\ \ insert}{kb\ size\ of\ \ vector} \times molar\ ratio\ of\ \frac{insert}{vector} = ng\ of\ \ insert$$

The approximate molar ratio of the insert and vector was kept as 2:1. Considering the average insert size to be 0.5 kb the ligation reaction was assembled as follows:

Components	**Stock concentration**	**Reaction concentration**	**Volume**
T_4 ligase buffer	10x	1x	1.0 µl
Digested vector	100 ng/µl	100 ng	1.0 µl
Digested DNA (insert)	50 ng/µl	33.33 ng	0.8 µl
T_4 DNA ligase	30 U/µl	12 U	0.4 µl
Water			6.8 µl

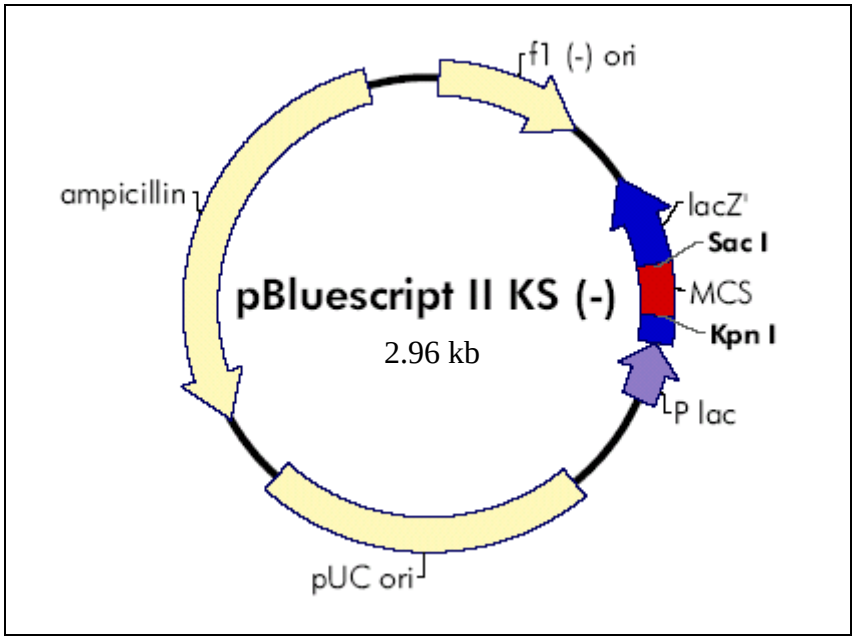

Figure 6.3: The pBluescript® II KS (-) phagemid vector

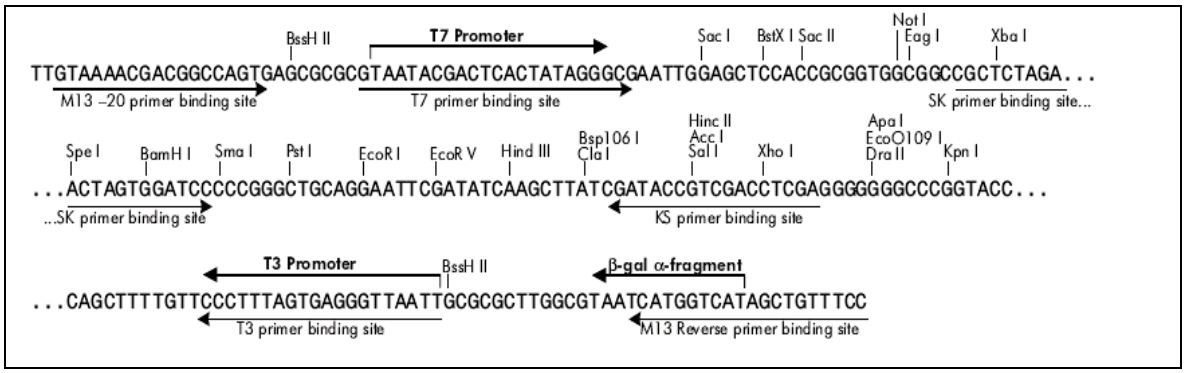

Figure 6.4: pBluescript® II KS (-) Multiple Cloning Site region

Table 6.1: Characteristic features of pBluescript® II KS (-)

Features	Nucleotide position
f1 (-) origin	21-327
β-galactosidase α-fragment	460–816
multiple cloning site	653–760
T7 promoter transcription initiation site	643
T3 promoter transcription initiation site	774
lac promoter	817–938
pUC origin of replication	1158–1825
ampicillin resistance (*bla*) ORF	1976–2833

For D4 and D6 separate ligation reactions were set up (the ligation products will be called L4 and L6 hereafter). The mixtures were incubated overnight at 14^{o}C. The success of ligation was confirmed by running 2.5 µl of the ligation product on 1.2% agarose gel. Since the size of the vector is 2.96 kb smearing from 2.96 kb up indicated successful ligation. Vector with no insert was run as a control.

6.2.2 Preparation of single stranded library

In order to perform the enrichment of microsatellite repeats, single stranded libraries for both batches of ligation product (L4 and L6) were generated by asymmetric PCR. The PCR reaction was carried out in 50 µl volume using a 10:1 molar excess for one of the primers. Universal primers, T3 and T7 (primer sites are present on pBluescript) were used with the T3 primers in excess. The following concentrations of the reagents were used: 1x Taq polymerase buffer, 1.5 mM of $MgCl_2$, 200 µM of dNTP, 50 ng T3 primer, 5 ng of T7 primer and 1U of Taq polymerase, 50 ng of ligation product and water to make the final volume of 50 µl. Thermocycling conditions were set as:

94^{o}C	4 min	
92^{o}C	1 min	} 32 cycles
50^{o}C	1 min	
72^{o}C	1 min	
4^{o}C	hold	

6.2.3 Enrichment of specific repeat sequence

Enrichment was carried out using two biotinylated tetranucleotide repeat probes viz. $(ATCT)_7$ and $(CTGT)_6$ which were captured by Streptavidin coated magnetic beads. For L4 and L6 separate reactions were set up.

6.2.3.1 Preparation of magnetic beads

i. 100 µl of magnetic bead suspension (MagnaSphere® Paramagnetic Particles, Promega) was pipetted into an Eppendorf tube.
ii. The tube was placed in a Magnetic Particle Concentrator (MPC) and left for 30 sec and the supernatant was removed.
iii. The tube was removed from MPC and the beads were resuspended in 100µl PBS buffer (0.15 M NaCl, 0.05 M $NaHP0_4$; pH 7.4) containing 0.1% BSA (Bovine Serum Albumin).
iv. Steps (ii) and (iii) were repeated once again to completely remove the preservative used for the magnetic beads.

6.2.3.2 Attaching biotinylated oligo to magnetic beads

Biotinylated oligos, $(ATCT)_7$ and $(CTGT)_6$, (from MWG Biotech) were used for this purpose. The oligos were attached to magnetic beads using the following steps.

i. 50 µl of the prepared magnetic beads were pipetted into each of two Eppendorfs.
ii. The beads were separated from the suspension buffer by placing the tubes in an MPC for 30s and then removing the supernatant.
iii. The tubes were removed from the MPC and the beads were resuspended in 100 µl 5x SSC (20x stock contains 3 M NaCl and 0.3 M sodium citrate, pH 7.0).
iv. 200 pmol (4 µl; 50 pmol/µl) of the biotinylated oligo was added to each tube (one type of oligo in each tube).
v. The mixtures were incubated at room temperature for 15 min with frequent gentle shaking.
vi. The tubes were placed on the MPC for 30s and then the supernatant was removed.
vii. The beads were resuspended in 100 µl 5x SSC.

viii. Steps vi and vii were repeated two more times to completely remove excess unbound oligo.

ix. After the last wash, the beads were resuspended in 35 µl 10x SSC.

x. The two bead-attached oligos were mixed at this stage in equal proportions.

6.2.3.3 Hybridization of DNA to oligo probes attached to magnetic bead

Thirty µl of the products of asymmetric PCRs were diluted with 35 µl of water giving a total volume of 65 µl. The diluted PCR product was denatured (to unfold any internal hybridization) for 10 min at 98^{o}C in a thermocycler. The tubes containing the denatured products were immediately placed on ice for about 1 min to avoid reannealing. Thirty five µl of the magnetic bead attached oligo was added at room temperature to the denatured DNA and the mixture was incubated with gentle agitation (frequent flicking) at room temp for 20 min. After incubation, the magnetic beads were first washed in low stringency washing solution (2x SSC; 0.5 ng/µl primer T3; 0.5 ng/µl primer T7) and then in high stringency solution (1x SSC; 0.5 ng/µl primer T3; 0.5 ng/µl primer T7). The low and high stringency washes were performed four times each. Each wash lasted 5 min at 30^{o}C. Supernatant from each wash was removed using an MPC and the beads were finally resuspended in high stringency solution. Supernatants from the high stringency washes were preserved in separate Eppendorfs for troubleshooting, if required.

6.2.3.4 Release of specific bound sequences from magnetic beads

To release the hybridized sequences from the magnetic beads, the bead suspension was incubated for 10 min at 75^{o}C. Immediately after incubation the suspensions were cooled down on ice for 1 min to avoid rehybridization. Beads were separated from suspension solution (now containing the released sequences) by placing on MPC. The supernatant

was preserved in a new Eppendorf. The supernatant from this final step and from the previous high stringency washes were desalted using the GFX^{TM} PCR DNA and Gel Band Purification Kit (Pharmacia) following the manufacturer's instructions.

6.2.4 Preparation of double stranded DNA

Symmetric PCR for each desalted product was carried out in 100 µl reaction volume. One reaction without any DNA was also included as a negative control. The PCR reaction was set up as: 1x Taq polymerase buffer, 1.5 mM $MgCl_2$, 200 µM dNTP, 70 ng T3 primer, 70 ng T7 primer, 2U of Taq polymerase, 20 µl of desalted product (template for PCR) and water to make a final volume of 100 µl. The thermocycling conditions were identical to those used for asymmetric PCR (section 6.2.2). The PCR product was checked on a 1.2% agarose gel. Smeary products observed on the gel indicated the success of PCR. The product of the PCR was desalted in the same way as mentioned above (section 6.2.3.4).

6.2.5 Double digestion

Double digestion of the amplified enriched sequences was carried out with *EcoR* I and *Hind* III to generate compatible ends for cloning into the *Eco*R I/*Hind* III prepared vector. The digestion was carried out sequentially, first in digestion buffer II (1x contains: 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM dithiothreitol) with *Hind* III and then in buffer III (1x contains: 100 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$ and 1mM dithiothreitol with *Eco*R I. The only difference between buffer II and III is that buffer III has a greater concentration of NaCl (100 mM). Thus, before digestion by *Eco*R I, NaCl was added in the digestion mixture to make the concentration 100 mM. The digestion reaction was set as: 1x RE buffer, 80 µl of the symmetric PCR product, enzymes in excess (≈ 50U each) and water to make 100 µl. The digestion

mixture was incubated at 37°C for about an hour and half. Immediately after digestion the product was desalted. Five µl of the desalted product was checked on agarose gel and quantified against the known concentration of φ174 *Hae* III. Figure 6.5 shows the result of double digestion. To separate the target DNA (i.e. the inserts) from the plasmid arm the entire digested product was run on 1% agarose gel. The desired band size was excised from the gel and purified by GFX™ PCR DNA and Gel Band Purification Kit (Pharmacia).

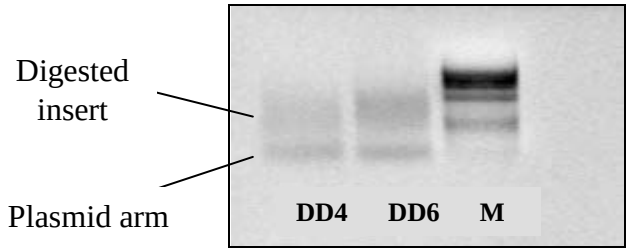

Figure 6.5: Result of double digestion observed on agarose gel. DD4 and DD6 stand for double digestion of products of 4 and 6 base recognition enzymes. M stands for φ174 *Hae* III size marker.

6.2.6 Ligation of the double digested product

The ligation was carried out into the *Eco*R I/ *Hind* III prepared pBluescript vector in the same way as was described in section 6.2.1.4. The only difference is that incubation was carried out overnight at 4°C (for sticky end cloning) in the fridge.

6.2.7 Transformation of competent bacterial cells and culture

One Shot® Top10 chemically competent *E. coli* (Invitrogen) were transformed with insert ligated vectors using heat shock. Fifty µl of supercompetent cells were transformed using 2 µl of ligated product. The mixture of cells and ligated vector was first incubated on ice for 30 min and then heat shocked at 42°C in a water bath for exactly 30 sec. The mixture was placed on ice immediately. Two hundred and fifty µl of

SOC medium (kept at room temperature) was added in each tube of transformed bacterial suspension. The tubes were shaken horizontally at 37^{o}C for 1hr and became ready for spreading on agar plates.

LB agar media for bacterial culture was prepared with 100 µg/ml of ampicillin and 40 mg/ml of X-GAL and was poured into a number of 35 ml petridishes. X-Gal was added for blue/white colony selection. X-Gal serves as the substrate for the enzyme beta-galactosidase of gene lacZ which is present in the pBluescript vector. When X-Gal is metabolized by beta galactosidase it produces bright blue coloured products and the colonies appear blue. The *Eco*R I and *Hind* III restriction sites reside in the lacZ gene. Because of this if the *Eco*R I-*Hind* III digested inserts successfully ligate between these sites, they disrupt the reading frame of the lacZ gene and destroy the functionality of beta-galactosidase enzyme. In such cases X-Gal is not metabolized and the colonies appear white. Therefore, white colonies indicate successful ligation of inserts into vector.

Initially, in each plate 10-20 µl of bacterial suspension was spread after diluting with 80-90 µl of LB broth. The plates were incubated at 37^{o}C overnight. These volumes of bacterial suspension resulted in too dense growth of colonies. Therefore, in all subsequent cultures 5 µl suspension diluted with 95 µl LB broth was used.

6.2.8 Picking positive colonies

Individual positive (white) colonies were hand picked using toothpicks and were dipped into 50 µl of LB broth media (containing 100 µg/ml of ampicillin) kept in individual wells of a microtitre plate. The plates were then left for incubation at 37^{o}C for 2 hours and then overnight at room temperature. For long term preservation of the colonies, 50 µl of 30% glycerine (with 100 µg/ml of ampicillin) was added to each individual well.

The microtitre plates were gently swirled to mix the glycerine with the bacterial suspension. The plates were sealed and preserved at -70^{o}C for future steps.

6.2.9 Transferring colonies onto nylon membrane

Bacterial suspensions preserved in microtitre plates were thawed. Hybond-N^{+} Nylon membranes (Amersham Biosciences) were cut in appropriate sizes so that one piece could accommodate colonies from two plates. The membranes were marked and labelled properly to indicate the placing of colonies from the microtitre plates and placed on rectangular agar plates. Colony transfer was performed using a 96-pin microtitre plate replicator. The replicator was sterilized by placing its pins into absolute alcohol and then by heating on flame. The sterilized comb was inserted into the suspension of the microtitre plate and then touched on the nylon membranes. The agar plates were covered and incubated at 37^{o}C overnight.

After overnight culture the nylon membranes with bacterial colonies were treated with denaturation buffer (87.66 g NaCl, 20 g NaOH in 1 litre solution with water) for 5 min and then twice with neutralization buffer (1.5 M NaCl, 0.5 M Trizma base; pH 7.5), each time for 5 min. Finally the membranes were washed in 2 x SSC solutions. Membranes were placed on Whatman papers to remove all the liquid and then dried in an oven at 80^{o}C for 2 hours. The membranes were kept at room temperature until used for hybridization with the labelled probe for detection of colonies containing repeats.

6.2.10 Hybridization with radioactively labelled probe

6.2.10.1 Radioactive labelling of probes

Although the enrichment was done only for tetranucleotide motifs, hybridization was performed with tetranucleotide as well as dinucleotide motifs viz. $(ATCT)_7$, $(CTGT)_6$,

$(GT)_{12}$ and $(CT)_{12}$. These oligos (from MWG Biotech) were end-labelled with radioactive ^{32}P. For end-labelling all the oligos were mixed in equal amount and the reaction was set up as follows: 1 µl (10 pmol) mixed oligo, 1.5 µl γ^{32}P-dATP (3 pmol), 1 µl of 10x T_4 buffer, 1 µl of T_4 enzyme and 5.5 µl of water. The reaction was incubated at 37^{o}C water bath for 2-3 hrs.

6.2.10.2 Hybridization

The nylon membranes with bacterial colonies fixed on them were treated with 0.1% SDS solution at 85^{o}C for 5 min and the bacterial cell-wall debris was removed by gently scraping the membranes with a piece of tissue paper. In a hybridization tube, 150 ml 4x SSC solution was placed and the membranes were rolled and placed in the tube. If several membranes were treated together they were alternated with nylon meshes. The tube was rotated in the hybridization oven at 60^{o}C for 15-20 min. The SSC solution was poured off and 50 ml of warm (62^{o}C) pre-hybridization buffer (prepared by adding 32.4 ml water, 4.5 ml 20x SSPE and 21 ml 20% SDS; pH 7.7) was added into the tube. Then the tube was left in the oven at 62^{o}C rotating for about 2-3 hrs for pre-hybridization incubation.

Before hybridization the radioactively labelled probes were left at 80^{o}C for about 5 min to denature any secondary structure. The prehybridization solution was removed from the hybridization tube and 15 ml hybridization buffer (prepared by adding 10 pmol probes in fresh prehybridization buffer) was poured in. The hybridization was carried out overnight at 53^{o}C.

Following overnight incubation the membranes were washed first with low stringency washing buffer (2x SSC with 0.5% SDS) followed by high stringency washing buffer (0.2x SSC with 0.5% SDS). The washing was performed twice with each buffer and

each wash lasted for 15 min at 60^{o}C. After every wash the intensity of radioactivity was measured: this reduced gradually. The membranes were blotted on tissue to remove excess liquid and wrapped with Saran cling film taking care that air bubbles were not trapped.

6.2.11 Identification of positive colonies with autoradiography

The membranes wrapped in Saran cling film were placed on autoradiography cassettes with the DNA sides up. X-ray film and an intensifier were placed on the membranes in a dark room. The cassettes were left in a -70^{o}C freezer overnight for exposure, wrapped in a black bag. The films were then developed. Positive colonies appeared as dark spots on the X-ray film (Figure 6.6).

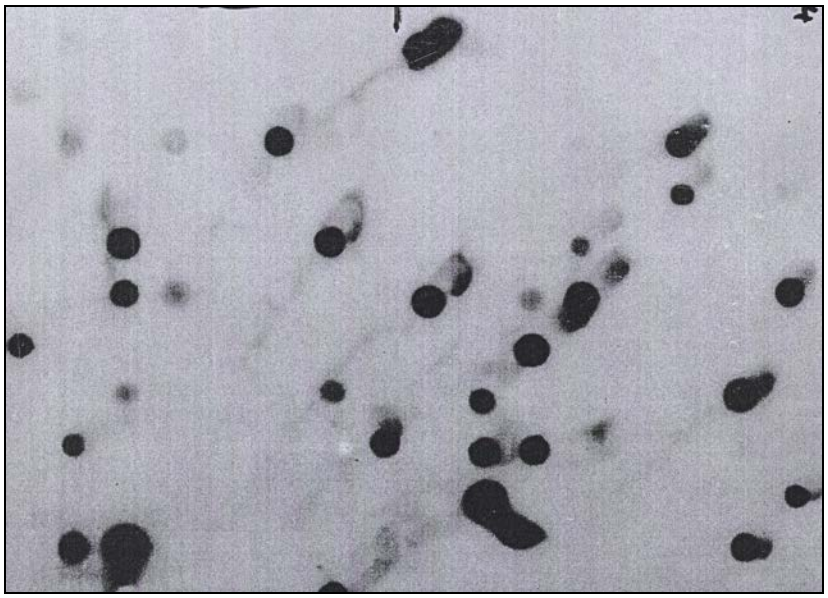

Fig 6.6:Image of autoradiography showing positive clones detected by radioactive probes

6.2.12 Isolation of plasmids

Each positive colony was located on the microtitre plate from their position on the X-ray film. New cultures were prepared from the positive colonies for extraction of

plasmids for sequencing. For each colony a 16 ml polypropylene tube was prepared with 3-4 ml of LB broth containing 100 µg/ml of ampicillin. The tubes were carefully labelled with the name of the clone. About 2 µl of the colony suspension was added into the respective tube and incubated overnight, shaking at 37^{o}C. Plasmids were isolated from each culture using the QIAprep® Spin Miniprep Kit (QIAGEN) following the manufacturer's protocol. The plasmid preps were quantified by spectrophotometry.

6.2.13 Sequencing

PCR for sequencing reactions was set up as follows: 1.6 µl Amersham DYEnamic ET Termination mix, 0.5 µl of M13F primer (10 pmol/µl), 2 µl of 500-800 ng plasmid prep, and 0.9 µl of double distilled water. Thermocycling setting was:

95^{o}C	30 sec	1 cycle
95^{o}C	20 sec	25 cycles
50^{o}C	15 sec	
60^{o}C	60 sec	
10^{o}C	hold	

DNA from the PCR mix was precipitated using the following steps:

a) In each tube 0.5 µl 1.5 M NaOAc /250 mM EDTA (pH 8.0) buffer was added.

b) Twenty µl of absolute alcohol was added in each tube and mixed using a vortex mixer and left at - 20^{o}C for 15 min.

c) The tubes were centrifuged at maximum speed (13,000 rpm) for 20 min. The supernatant from each tube was carefully removed.

d) One hundred and eighty µl 70% alcohol was added to each tube and then centrifuged at maximum speed for 5 min. The supernatant was removed.

e) The samples were dried. Two µl formamide load dye was added, mixed well and left for 30 min for resuspension before sequencing.

Sequencing was performed with an ABI^{TM}PRISM 377 automated sequencer. The preparation of plates and denaturing polyacrylamide gel was as described in Chapter 2

(section 2.6.1). Sequencing data was processed and analysed by *ABI Prism DNA Sequencing Analysis Software*. Microsatellite repeats in each sequence were searched for using the web-based programme 'Tandem Repeats Finder ver. 3.21' (Benson 1999). The sequences were compared to each other applying "Local Blast" using the programme BioEdit (Hall 1999). To find out if the already available rohu microsatellite sequences in the GenBank database matches with the any one resulting from present isolation, all the available online sequences were copied to BioEdit and subject to Local Blast.

6.2.14 Primer design

Primers were designed by the programme QUICKPRI PrimerSelect in the DNASTAR (ver. 3.04) software package. Initially the default primer characteristics set by the programme were used for designing primers.

Primer length: *18 bp to 24 bp*
3′ end pentamer stability*: 8.5 –kc/M*
Unique 3′ sequence*: 7 bp*
Accept dimer duplex: *2 bp*
Accept hairpin duplex: *2 bp*

If the default setting failed to produce a satisfactory result then the conditions were relaxed by changing the primer length and increasing the values for dimer duplex and hairpin duplex. Primer pairs were selected in a way so that the melting temperatures (T_m) of both primers in a pair were similar (preferably not more than 5^oC difference). Primers with too low T_m (below 50^oC) and too low percent GC content (below 35%) were preferentially avoided. Primer pair stability score was set to be high.

6.2.15 Characterisation of microsatellites

The microsatellites were characterized for the following features:

- Amplification conditions for each locus using gradient PCR
- Polymorphism of each locus
- Allele frequency distribution at each locus
- Polymorphic Information Content (PIC) and heterozygosity of each locus

The success of PCR amplification was checked on 1.2% agarose gel. The best amplification-condition (i.e. annealing temperature and $MgCl_2$ concentration) was detected by comparing the products of gradient PCR on agarose gel. Polymorphism of the loci was checked by genotyping 8-16 unrelated individuals for each locus. These individuals came from a mixed sample of rohu collected from three different river systems (hence presumably three different wild populations) in Bangladesh. (Unfortunately, while bringing the sample from Bangladesh, labels on the tube were accidentally erased through spillage of alcohol and it was not possible to detect which individual originated from which river system). The rationale of using fish originating from heterogeneous wild sources was to obtain a broader perspective about the polymorphism of the loci. Due to very small sample size Hardy-Weinberg Equilibrium test could not be performed reliably.

A Beckman-Coulter CEQ^{TM} 8000T capillary sequencer was used for fragment analysis (details in Chapter 2, section 2.6.2). M13 tailed primers and fluorescent M13 dye labelled primers were used for attaching fluorescence to the PCR products. Genotype data was further analysed by the programme CERVUS for calculating allele frequency, heterozygosity and PIC, for performing HWE test in the sample and for estimation of null allele frequency.

6.3 Results

The success of microsatellite isolation from the rohu genomic library is summarized in Table 6.2. From the culture of transformed super-competent *E. coli*, 1,248 positive (white) colonies were lifted and subject to autoradiography to detect microsatellite repeats with four different radioactive probes viz. $(GT)_n$, $(CT)_n$, $(ATCT)_n$ and $(GTCT)_n$.

Table 6.2: Summary result of microsatellite isolation from rohu (*Labeo rohita*)

Features	**Number and percent**
White colonies screened with autoradigraphy	**1248**
➢ CI_4*	672
➢ CI_6*	576
Colonies giving positive signal	**148 (11.85%)**
➢ CI_4	68 (10.12%)
➢ CI_6	80 (13.89%)
Colonies sequenced	**114**
➢ CI_4	54
➢ CI_6	60
Number of readable sequences	**101**
Sequences having microsatellite repeats	**101 (100%)**
Number of sequences used for primer design	**50 (49.5%)**
Sequences that could not be used for primer design	**51 (51.5%)**
➢ Due to lack of flanking sequence	37 (72.54%)
➢ Due to too short repeat length (<5 units)	11 (21.56%)
➢ No suitable primer pairs found	3 (5.88%)
Number of primer pairs designed	**52**
Primer pairs successfully amplifying product	**48**
Number of polymorphic loci	**36 (75%)**

* CI_4 stands for colonies with inserts produced by digestion with 4 base-recognition restriction enzymes
CI_6 stands for colonies with inserts produced by digestion with 6 base-recognition restriction enzymes

One hundred and forty eight colonies (11.85%) produced positive signals. Although more colonies with inserts restricted with 4-base-recognition enzymes (denoted as CI_4 in Table 6.2) were screened compared to colonies having inserts digested with 6-base-recognition enzyme (CI_6), less of the CI_4 colonies produced positive signals in autoradiography (10.12%, compared to 13.89% for CI_6).

Colonies giving reasonably strong signals (114) were sequenced from which 101 colonies gave readable sequence. All the readable sequences contained microsatellite repeats although in a number of cases (about 15%) the repeat array was short (up to 5 repeat units). The repeat numbers in the sequences varied from 3 to 49. Figure 6.7 shows the frequency distribution of the number of repeat units of the microsatellite arrays. The greatest percentage (27.35%) of the microsatellites showed repeat units in the range of 11-15 followed by 6-10 (23.93%). Approximately 5% of the loci had repeat numbers above 30.

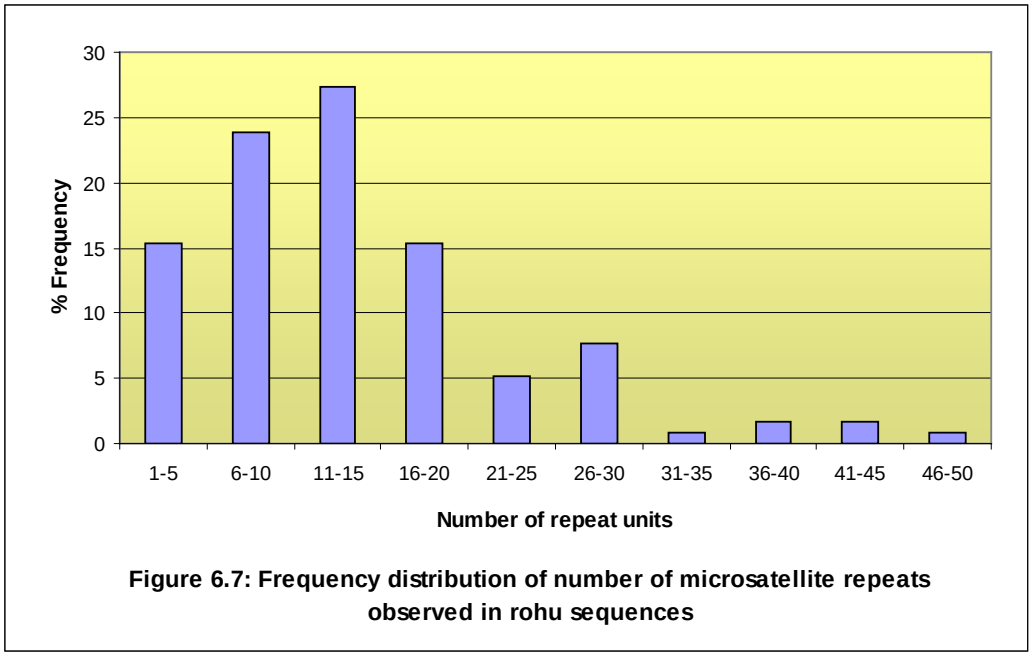

Figure 6.7: Frequency distribution of number of microsatellite repeats observed in rohu sequences

From 101 readable sequences primers could be designed from 50 sequences and a total of 52 pairs of primers were designed (2 sequences contained 2 separate microsatellite arrays each and primers were designed for both). Failure to design primers from the other sequences (51.5% of sequences) was a result of two major reasons: lack of flanking sequence at either or both ends of the microsatellite array (about 73% of the cases) or short repeat number, generally below 5 repeat units (about 22% of the cases). In only a few cases (about 6%) it was not possible to design compatible primer pairs from the available flanking region. Forty eight primer pairs successfully amplified products in PCR. Of these 36 loci (75%) were found to be polymorphic, that is showed more than one alleles. Four loci failed to amplify in PCR reaction. Attempts were made to redesign the primers for these loci, but in three cases no new compatible pairs were found and in one case even after redesigning the primers the loci failed to amplify probably because of poor primer stability.

6.3.1 Microsatellite motifs

Table 6.3 summarizes the types of repeat motifs found from microsatellite screening of rohu sequences. Even though the enrichment was carried out using tetranucleotide motif probes, only about 18% (14.53% for GTCT/CAGA and 3.42% for ATCT/TAGA) of the sequences contained those motifs. Some other tetranucleotide motifs (GTTT, ATCC) were also observed which have been placed in the 'miscellaneous' group in the first part of Table 6.3. The predominant types of microsatellite motifs were dinucleotide repeats, especially GT/CA which appeared in 67.52% of the sequences. CT/GA repeats appeared in 10.26% of the microsatellite arrays. Only one trinucleotide repeat was found. By form, the microsatellite repeats were grouped into three types: (i) pure or perfect repeats (50.43%), (ii) Interrupted or imperfect repeats (41.03%) and (iii)

compound repeats, (8.55%) where more than one type of repeat motif appeared consecutively or in close proximity.

Table 6.3: Result summary of the repeat motifs and type of microsatellite repeats screened in rohu

Total number of microsatellite repeat arrays (in 101 sequences)	117
Repeat motifs and their frequency	
GT/CA	67.52%
CT/GA	10.26%
GTCT/CAGA	14.53%
ATCT/TAGA	3.42%
Miscellaneous	4.27%
Type of repeats by nucleotide string	
Dinucleotide	78.63%
Trinucleotide	0.85%
Tetranucleotide	20.51%
Type of repeats by form	
Pure/ perfect	50.43%
Interrupted/imperfect	41.03%
Compound	8.55%

6.3.2 Primer design

Out of 52 microsatellite loci for which primer pairs could be designed, 48 amplified successfully. Table 6.4 reports the primer information of these 48 loci along with the melting and annealing temperatures for the primers. The length of the designed primers ranged between 18-27 bases and the percent GC content from 20-68%. Generally a GC content of 40-60% is considered to be the best for primer stability. About 16% of the primers had GC content less than 40% while about 11% of the primers had greater than 60% GC content. Most of the primer pairs amplified the respective loci over a wide range of annealing temperatures (T_a). The workable range of T_a reported in Table 6.4 indicates the broad range of temperatures where a primer pair amplified product whereas the best T_a range indicates the range where the amplification was the strongest.

Table 6.4: Primer information for the rohu microsatellite loci that amplified successfully. All loci were amplified with 1.5mM $MgCl_2$.

Locus name		Primer sequence	Primer length	% GC content	T_m (^oC)	Workable T_a range (^oC)	Best T_a range (^oC)
*Lro*1	*F:**	TGTCATTATACAAATTTGTTTCCTGA	26	26.92	55.3	49-59	53-59
	*R:**	TGTGTATATGAGTGTTTGTGTCTGTCT	27	37.04	60.4		
*Lro*2	*F:*	TCGACCATGCTTGTCTTTTGTTTA	24	37.50	57.6	49-59	49-59
	R:	CATGGAAGCATCACTTTGTTATCG	24	41.67	59.3		
*Lro*3	*F:*	TTAGCCGCTTCAGTTCATTCA	21	42.86	55.9	49-59	49-59
	R:	TTAGATCCCCACCGCCTTAT	20	50.00	57.3		
*Lro*4	*F:*	AGGCCAAACGCTCACTCAG	19	57.89	58.8	51-62	51-62
	R:	GACACCCGCAGAACCTCACT	20	60.00	61.4		
*Lro*5	*F:*	TGACGCCGACGTGAATGTCAC	21	57.14	61.8	57-63	57-63
	R:	CTGCTGATGACCGTCAACATGAC	23	52.17	62.4		
*Lro*6	*F:*	CTCTCCACCCTCCATACCACGAC	23	60.87	66.0	57-66	63-66
	R:	TGCCCTGCATTTCTCCCATCC	21	57.14	61.8		
*Lro*7	*F:*	GGTGGATTAGCGTGGGTGACT	21	57.14	61.8	51-62	57-62
	R:	TTTTGTGCTTGAATCGGACTTG	22	40.91	56.5		
*Lro*8	*F:*	TGTTCTTATGCTAACTCAGAGTGACG	26	42.31	61.6	51-62	51-62
	R:	ATCTTCATCCATTGCATAAAGACAG	25	36.00	58.1		
*Lro*9	*F:*	GCCAGCGGCCCTGCACAGTAA	21	66.67	65.7	59-66	59-66
	R:	TGTCCGTTTCAGGCTAAAGCAGTCTG	26	50.00	64.8		
*Lro*10	*F:*	AGGCAGTGAGGATAATTGTGCTCT	24	45.83	61.0	55-63	55-63
	R:	CTGGCCTGCGAGTGACCAT	19	63.16	61.0		
*Lro*11	*F:*	TTTTTCGCAAAGTTTGCATGGCTGAA	26	38.46	60.1	57-66	57-66
	R:	CTGCCGAAAACACTGAACACGGACAG	26	53.85	66.4		
*Lro*12	*F:*	CAGCGCTGGAACGACACCA	19	63.16	61.0	55-63	55-59
	R:	TGCTGCGGGTCATTAGTATTCATC	24	45.83	61.0		
*Lro*14	*F:*	AGGCGCAGTATTGTTTAGGCTCTGTG	26	50.00	64.8	59-66	59-66
	R:	GCCCTCCTCAGTCATAAAACCCAATC	26	50.00	64.8		
*Lro*15	*F:*	CGATTGACTGCTATTGTATTTTTA	24	29.17	54.2	47-57	47-57
	R:	TAGACGATTATGTGGATGGA	20	40.00	53.2		
*Lro*16	*F:*	CTTCAAAGTCCTAGTGCTCTACACG	25	48.00	63.0	55-66	55-66
	R:	CTGCCTGCCTCTACTGCTGTCTA	23	56.52	64.2		
*Lro*17	*F:*	GGGGCCATTCAGACCAGCTTATCTAT	26	50.00	64.8	55-66	55-66
	R:	GACCCATTGCTGAAATCAACCTGAGT	26	46.15	63.2		
*Lro*18	*F:*	TTTGCAGGGTGACGTTACACAGTA	24	45.83	61.0	53-62	59-61
	R:	TTTTGATTTACGTACCCCGTTTCA	24	37.50	57.6		
*Lro*19	*F:*	AGGTTGATCTCCCACTCATGTGATTA	26	42.31	61.6	53-62	53-63
	R:	AAAGGTTATCTGTGGCGTATGAAGGT	26	42.31	61.6		
*Lro*20	*F:*	TCATCTGAACCCTTTTATTTG	21	33.33	52.0	48-59	52-58
	R:	CGTTTACCTGCGGAGACA	18	55.56	56.0		
*Lro*21	*F:*	TTAAAAGCAAAGGGTGGTG	19	42.11	52.4	48-55	48-55
	R:	TCAGGGGAAAAAGCAACA	18	44.44	51.4		
*Lro*22[1]	*F:*	CTAGTCCCCATGAGTCTGTGT	21	52.38	59.8	53-62	53-63
	R:	GTGTGCGAGTGCGAGTGA	18	61.11	58.2		
*Lro*23[1]	*F:*	ACACTCATACTCACTCGCACTCG	23	52.17	62.4	55-66	57-61
	R:	CTTCAGCCGCCTGTCAGTAAT	21	52.38	59.8		
*Lro*25	*F:*	CGGTGAATTTGCAGTGATGTGT	22	45.45	58.4	53-62	53-63
	R:	CAACTACTGCAACCTGAGAACG	22	50.00	60.3		
*Lro*26	*F:*	AGATCATTGCTGGGGAGTGTTTAT	24	41.67	59.3	53-62	57-62
	R:	GACCTGCCTGTGCCATCTGTA	21	57.14	61.8		

Table 6.4 continued…

Locus name		Primer sequence	Primer length	% GC content	T_m (°C)	Workable T_a range (°C)	Best T_a range (°C)
*Lro*27	*F:*	GTGTGAGTTGATGATGGTTTAGTA	24	37.50	57.6	62	62
	R:	AGAGCAGGTGGGTTTGAAT	19	47.37	54.5		
*Lro*29	*F:*	AATGGCGCTTGAACAGAATC	20	45.00	55.3	51-62	51-62
	R:	GCGACCGCTTTAATGAGACGA	21	52.38	59.8		
*Lro*30[2]	*F:*	TTATCACCGTCAAACCACAC	20	45.00	55.3	51-62	56-58
	R:	TGCTGCTCGCCACTGCTA	18	61.11	58.2		
*Lro*31[2]	*F:*	CATAATAGCAGTGGCGAGCAG	21	52.38	59.8	53-62	59-62
	R:	AACCACCAGCACACCTTTCAC	21	52.38	59.8		
*Lro*32	*F:*	ACCCTCTTTGTTTTGGCTCTC	21	47.62	57.9	52-60	55-57
	R:	TCTCTTACCCTGTTTCTCTGT	21	42.86	55.9		
*Lro*33	*F:*	AGTAACTTATGTATCGGCTTCAAC	24	37.50	57.6	50-62	57-60
	R:	TTACGGACAAAATCACTTACATC	23	34.78	55.3		
*Lro*34	*F:*	GCGTTGGTCTGGGGTGAA	18	61.11	58.2	53-62	57-61
	R:	AGATAGACGGACGGGTTTACGA	22	50.00	60.3		
*Lro*35	*F:*	CGCAGTGGGATACGCATTACAT	22	50.00	60.3	53-62	53-62
	R:	TCGGCCGCAGTGAGCATC	18	66.67	60.5		
*Lro*36	*F:*	GTTGTGGCTGGATTTATTACT	21	38.10	54.0	51-62	52-59
	R:	GGTGAGCCATTGACAGGT	18	55.56	56.0		
*Lro*37	*F:*	ATGTTGTGGTCATCATGTAAATC	23	34.80	55.3	57-59	57
	R:	CAGTTTCCTCCCTTCATAGTTT	22	40.90	56.5		
*Lro*38	*F:*	TAGGTTTAGGGTTAGGGGATAG	22	45.45	58.4	59	59
	R:	TTGTAATAGTTGCAAATGAGTC	22	31.82	52.8		
*Lro*39	*F:*	TCACTGAGCACAGGAAGGCAGGAATG	26	53.85	66.4	57-66	61-66
	R:	ATTGGTGCCGACGGAACAGGAAGTCT	26	53.85	66.4		
*Lro*40	*F:*	AATAAAAGGTGTTCGGTGTCTGTA	24	37.50	57.6	52-61	52-61
	R:	GAGCTCTCCGATCGATAGTCAAC	23	52.17	62.4		
*Lro*41	*F:*	GACTTCAGCTTCTCCACTCAA	21	47.62	57.9	52-60	52-59
	R:	CTGCGTTTAACCAATCACAA	20	40.00	53.2		
*Lro*42	*F:*	CCGTTCATGCGGTGGAGGTCT	21	61.90	63.7	58-66	58-66
	R:	CTGAAGCTCAGGCTGTCCATCTCC	24	58.33	66.1		
*Lro*43	*F:*	TCTCTGCGCCTGTCTACCT	19	57.89	58.8	52-56	52-54
	R:	TGTTTATTAAAGCACTTTCCTCAT	24	29.17	54.2		
*Lro*44	*F:*	TCAGTCTTTAAGCGTGTGGAGTGC	24	50.00	62.7	57-64	58-63
	R:	ATGGGAACGAGGAGAGGACGAA	22	54.55	62.1		
*Lro*45	*F:*	TGGTTACTGTAGCAAAACTGGTGT	24	41.67	59.3	54-57	55
	R:	TGACCTTTAGGGGTACAACAGC	22	50.00	60.3		
*Lro*46	*F:*	CTGCGACGGAAAGACGAT	18	55.56	56.0	52-60	52-60
	R:	AGAGCAGCAGGTGTATTATGAAGT	24	41.67	59.3		
*Lro*47	*F:*	CTGCCACAGGAGTGTGATCC	20	60.00	61.4	55-66	58-65
	R:	TATGCGGTCTATGGCTCTGTGTTC	24	50.00	62.7		
*Lro*48	*F:*	AGGGAGGCGGGGTGGAAGA	19	68.42	63.1	55-65	55-65
	R:	CACATGACCTAAAGCCGGATTTTC	24	45.83	61.0		
*Lro*49	*F:*	TTGCTCGACACGCTGAAA	18	50.00	53.7	52-60	55-59
	R:	CACCCTTACTCGCTGATGTC	20	55.00	59.4		
*Lro*50	*F:*	CCTAAACGCTGCCTTGTAAAGAAGC	25	48.00	63.0	57-65	57-65
	R:	ATGCGAGGCCACGGTGTCAG	20	65.00	63.5		
*Lro*52	*F:*	TTGTGCGTATCTGTGAGCTGATGAGT	26	46.2	63.2	58-66	58-66
	R:	AACAAGAATTTATCGCCTCGCTCTCC	26	46.15	63.2		

* F stands for forward primer and R stands for reverse primer

[1] *Lro*22 and *Lro*23 were developed from different repeat arrays in a single sequence

[2] *Lro*30 and *Lro*31 were developed from different repeat arrays in a single sequence

6.3.3 Characterisation of rohu microsatellite loci

Thirty six microsatellite loci were found to be polymorphic. Most of the loci (28) contained pure or interrupted GT repeats, 7 loci had compound repeats and only one locus contained the tetranucleotide GTCT repeat for which enrichment was performed. Table 6.5 presents the repeat motifs, level of polymorphism (in term of number of alleles, heterozygosity, polymorphic information content), size range, prediction for null alleles and the GenBank accession number of individual loci. Hardy-Weinberg equilibrium test could not be performed for most the loci because of the low frequency of some alleles in a locus and also because of the low number of individuals scored. That is why the result of HWE has not been shown in Table 6.5. Figure 6.8 shows the distribution of the number of alleles per microsatellite locus. The loci were arbitrarily classified in three categories based on their level of polymorphism observed by scoring 8-16 individuals:

(i) Low polymorphism (2-4 alleles): 20 loci

(ii) Medium polymorphism (5-6 alleles): 7 loci

(iii) High polymorphism (>6 alleles): 9 loci

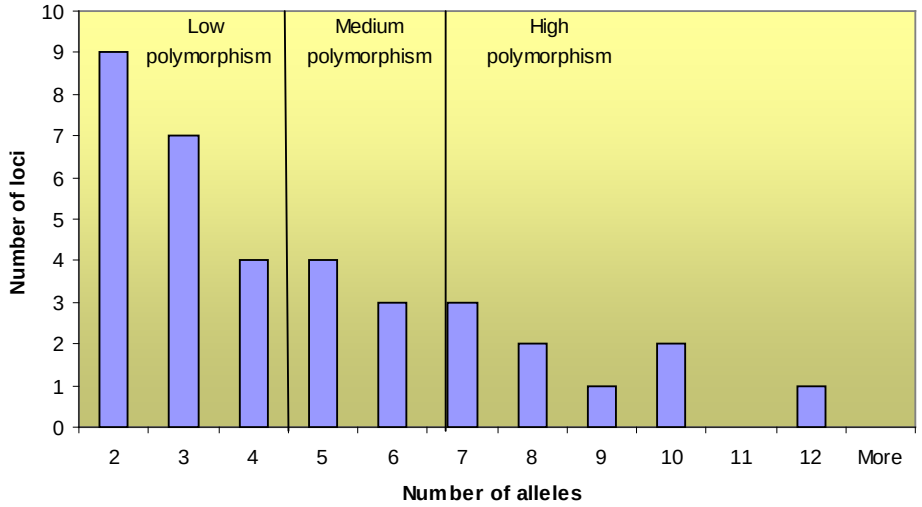

Figure 6.8: Frequency distribution of the number of alleles across loci

The expected heterozygosity of the loci varied from 0.083 to 0.935. Among these 50% of loci (total 18) showed H_e greater than 0.5 (Table 6.5). The Polymorphic Information Content (PIC) values varied from 0.077 to 0.872. Five loci namely, *Lro*20, *Lro*33, *Lro*50, *Lro*43 and *Lro*04 showed a large positive null allele prediction. However, the null allele prediction in this context should be taken with caution as the sample size was very small. Figure 6.9 shows the allele frequency distribution in different loci.

Table 6.5: Characterization of polymorphic rohu (*Labeo rohita*) microsatellite loci

Locus	Clone name	Repeat motif	A	N	Size range	H_o	H_e	PIC	Null freq	Genbank accession
*Lro*01*	T4P1A7	$(CA)_2GA(CA)_3T_2(CA)_{14}$	5	15	244-262	0.400	0.453	0.411	0.0132	AM184128
*Lro*02*	T4P4C1	$(GT)_7ACT_2(GT)_{17}AT(GT)_4$	10	15	205-227	1.000	0.885	0.841	-0.0836	AM184129
*Lro*03*	T4P4C5	$(GT)_8T_2$ $(GT)_3$	3	15	220-234	0.267	0.246	0.221	-0.065	AM184130
*Lro*04	T4P4D5	$(CA)_{19}$	6	12	209-223	0.583	0.833	0.768	0.1565	AM184131
*Lro*05	T4P4D9	$(GTTT)_3AT_3AT(GT)_{15}AT(GT)_{10}$	12	9	182-231	0.889	0.935	0.872	0	AM184132
*Lro*06	T4P4E10	$(GT)_{29}$	7	7	196-218	1.000	0.857	0.768	0	AM184133
*Lro*10	T4P4F2	$(GT)_{10}$	2	12	135-139	0.083	0.083	0.077	-0.0122	AM184134
*Lro*11*	T4P4G5	$(GT)_9$	2	16	182-184	0.188	0.175	0.155	-0.0423	AM184135
*Lro*12	T4P4G9	$(CA)_3CT(CA)_{15}$	4	12	111-125	0.500	0.562	0.481	0.0317	AM184136
*Lro*14	T4P4H8	$(GT)_{12}$	3	8	175-181	0.375	0.492	0.398	0	AM184163
*Lro*15	T4P5F5	$(GT)_{15}$	4	11	182-188	0.545	0.571	0.471	0.0076	AM184137
*Lro*17	T4P6C7	$(CA)_{15}$	4	12	151-159	0.667	0.714	0.62	-0.0032	AM184138
*Lro*20	T4P6G12	$(GT)_8$	2	12	154-156	0	0.159	0.141	0.8686	AM184139
*Lro*21	T4P6H10	$(GT)_{11}$ $TT(GT)_4$	3	12	198-209	0.417	0.518	0.408	0.0823	AM184140
*Lro*22[1]	T4P6H6	$(CA)_{15}$	2	6	175-179	0.333	0.303	0.239	0	AM184141
*Lro*23[1]	T4P6H6b	$(CA)_5TCTA(CA)_6CGCT(CA)_5$	2	7	177-187	0.429	0.363	0.28	0	AM184142
*Lro*25*	T4P7E3	$(GT)_{14}$	6	15	206-224	0.667	0.772	0.709	0.0516	AM184143
*Lro*26*	T6P1C12	$(GT)_2C(GT)_{17}AT(GT)_5$	5	15	326-340	0.733	0.685	0.614	-0.0595	AM184144
*Lro*29	T6P2A10	$(GTCT)_8$	3	8	198-218	0.375	0.342	0.294	0	AM184145
*Lro*30[2]	T6P2A4a	$(GT)_{11}$	3	12	133-139	0.250	0.236	0.212	-0.0597	AM184146
*Lro*31*[2]	T6P2A4b	$(GT)_{22}$	10	16	183-219	0.875	0.792	0.743	-0.0849	AM184147
*Lro*32	T6P2C5a	$(GT)_{17}$	8	12	150-164	0.750	0.841	0.782	0.0297	AM184148
*Lro*33	T6P2D7	$(GT)_{18}$	4	14	150-158	0.214	0.619	0.528	0.4721	AM184149
*Lro*34*	T6P4B2	$(AT)_6(GTAT)_2(AT)_2(GT)_{18}$	7	16	143-169	0.938	0.794	0.739	-0.1061	AM184150

Table 6.5 continued….

Locus	Clone name	Repeat motif	A	N	Size range	H_o	H_e	PIC	Null freq	Genbank accession
*Lro*35	T6P4B9	$(CA)_{11}GA(CA)_2$	3	8	142-148	0.250	0.242	0.215	0	AM184151
*Lro*36	T6P4D4	$(GT)_{13}$	5	11	293-308	0.545	0.641	0.541	0.0615	AM184152
*Lro*37	T6P4D6	$(CA)_3CG(CA)_9(CT)_2T(CT)_3GA(CA)_4$	3	11	326-338	0.364	0.437	0.360	0.0588	AM184153
*Lro*39	T6P4E11	$(GT)_{14}$	5	10	250-260	0.700	0.674	0.603	-0.0347	AM184154
*Lro*40	T6P4G3	$(CA)_5(TA)_2(CA)_5$	2	12	235-237	0.083	0.083	0.077	-0.0122	AM184155
*Lro*41	T6P4H9	$(CTT)_2(GT)_8(CT)_7$	2	12	171-173	0.250	0.228	0.195	-0.062	AM184156
*Lro*42	T6P5A6	$(GTT)_4$…..$(GT)_{17}$	7	12	159-175	0.750	0.685	0.608	-0.0859	AM184157
*Lro*43	T6P5B11	$(GA)_{39}G(CT)_2(GT)_8$	8	10	164-223	0.600	0.747	0.689	0.1268	AM184158
*Lro*44	T6P5C4	$(GT)_{16}$	6	12	178-190	0.917	0.710	0.643	-0.1841	AM184162
*Lro*47	T6P6A9	$(GT)_{10}T_2AT(GT)_5(CTGT)_2(CT)_2(GT)_4$	2	12	233-235	0.167	0.159	0.141	-0.0358	AM184159
*Lro*49	T6P6F5	$(GT)_5AT(GT)_4AT(GT)_4AT(GT)_{14}$	9	12	189-216	0.917	0.833	0.774	-0.0725	AM184160
*Lro*50	T6P6G11	$(GT)_4AT(GT)_9$	2	12	138-140	0.083	0.228	0.195	0.4439	AM184161

A = Number of alleles
N = Number of individuals scored
H_o = Observed heterozygosity
H_e = Expected heterozygosity
PIC= Polymorphic Information Content
* These loci were characterised by Sanjeev John, as part of his MSc research project
[1] *Lro*22 and *Lro*23 were developed from different repeat array in a single sequence
[2] *Lro*30 and *Lro*31 were developed from different repeat array in a single sequence

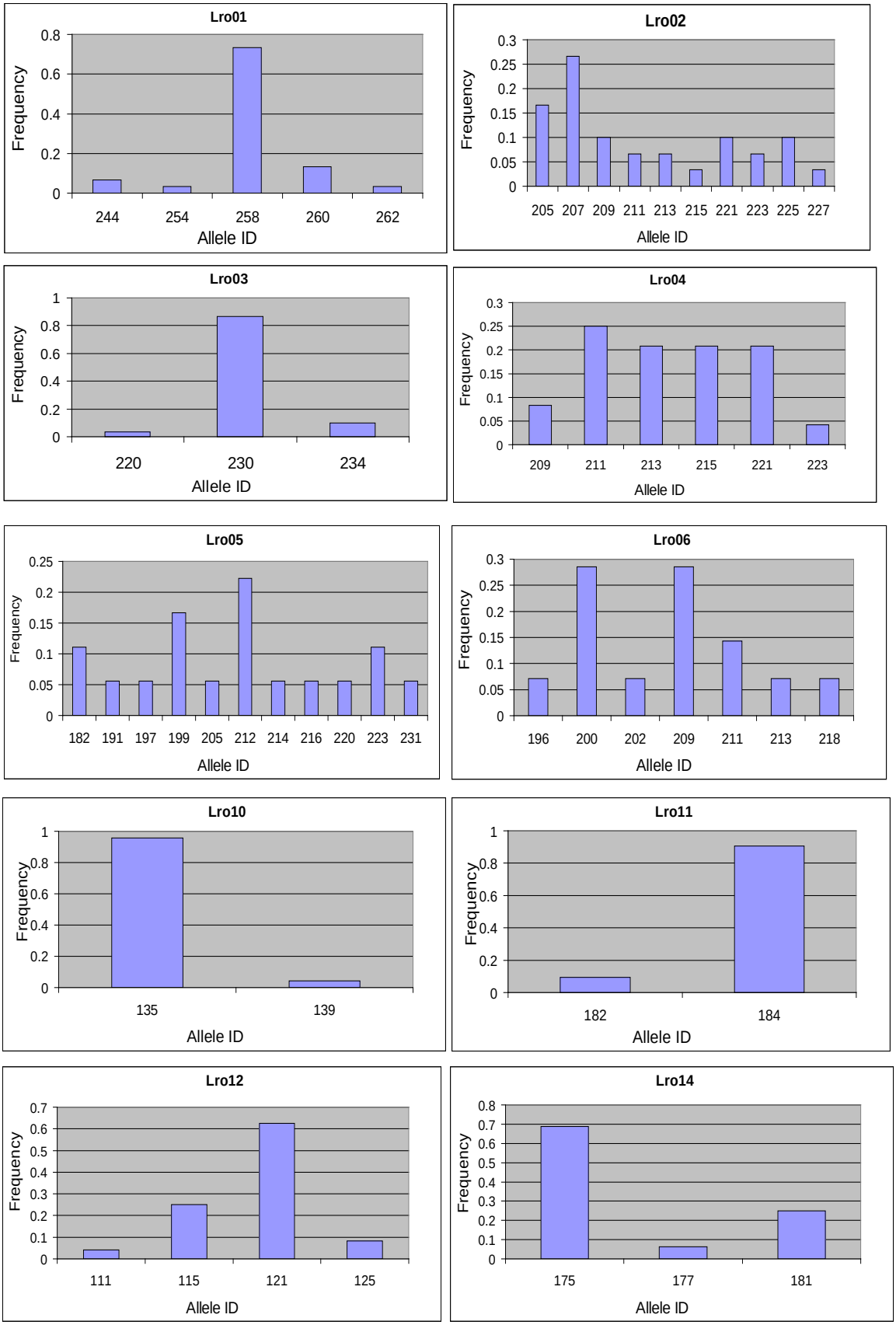

Figure 6.9: Graphs showing the allele frequencies across different Rohu microsatellite loci

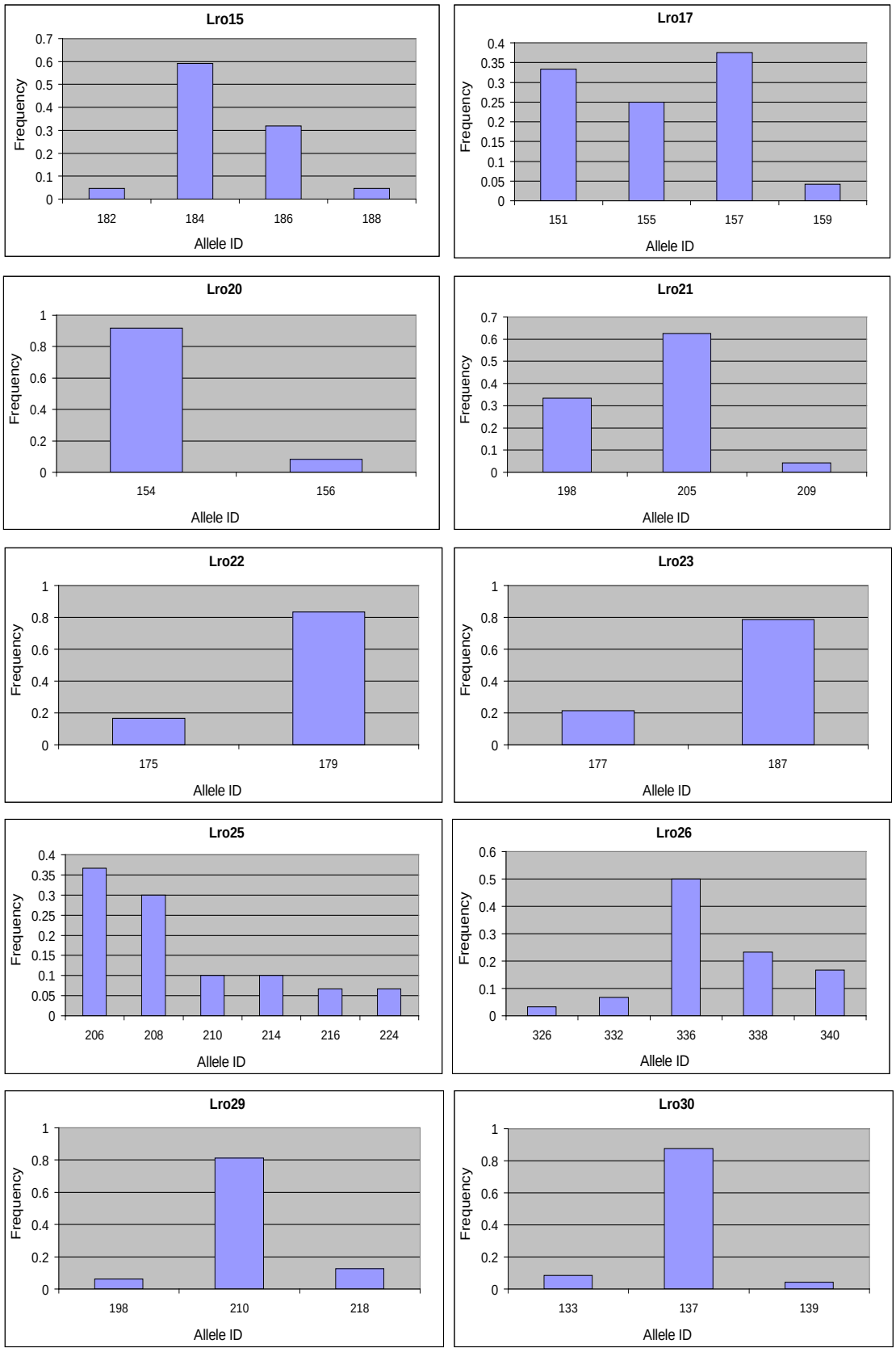

Figure 6.9 continued ……

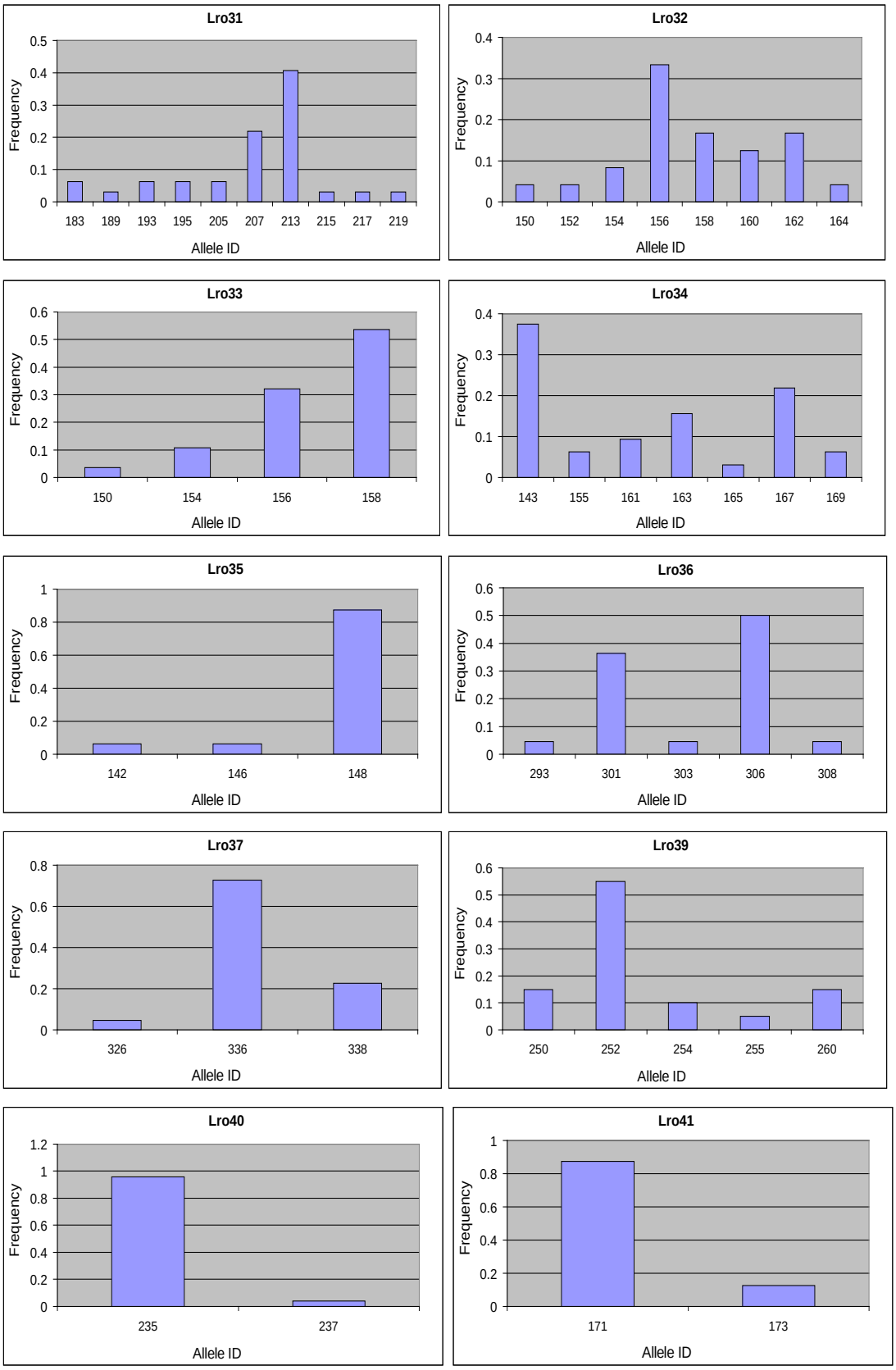

Figure 6.9 continued ……

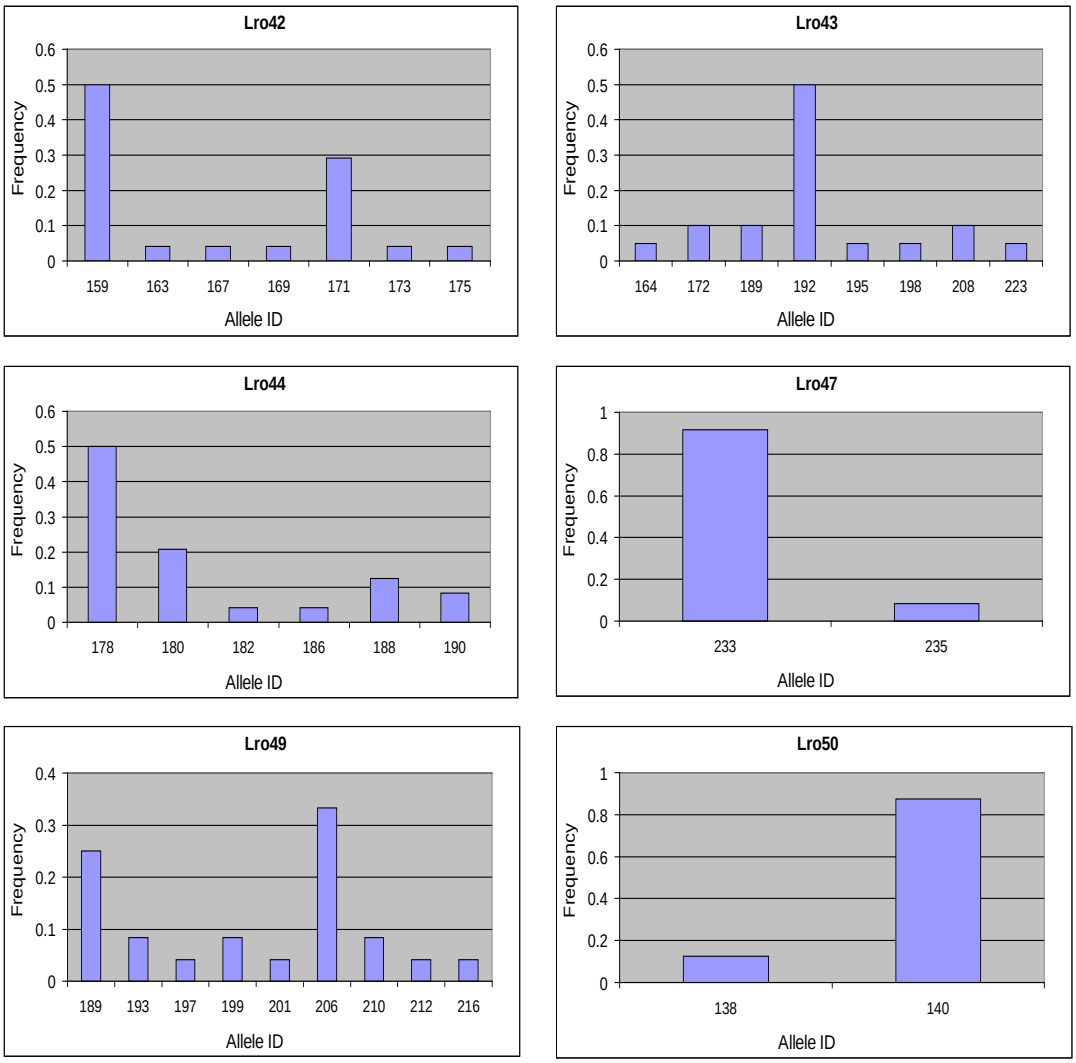

Figure 6.9 continued ……

6.4 Discussion

The present chapter describes the isolation of microsatellites from a partial genomic library of rohu using a selective hybridization enrichment protocol. Primer pairs have been designed for 52 microsatellite loci from which 36 loci were found to be polymorphic when genotyped on 8-16 unrelated individuals.

Enrichment was performed for tetranucleotide repeats ATCT and GTCT. Although microsatellites with dinucleotide repeats are the most abundant (excluding the mononucleotide) types in vertebrate genomes, scoring of alleles at these loci can be problematic because of stutter bands. Stutter bands are produced during the PCR mainly as a result of replication slippage, generally masking alleles differing by two base pairs in size (Naish *et al.* 1998). In contrast, microsatellites with tetranucleotide repeats can be easily scored as PCR amplification produces few if any extra products. This was the reason for performing enrichment with tetranucleotide repeats. In vertebrates ATCT repeats have been reported to be the most prevalent type of tetranucleotide motif, followed by GTCT (reviewed in Zane *et al.* 2002). Hence, these two motifs were used for enrichment.

The efficiency of an enrichment protocol depends on the type of repeat motif used and also on the species from which the microsatellite is being isolated. Maguire (2001) reviewed the success of enrichment with the selective hybridization approach used by several groups of workers. In the reported studies 20% -100% of the positive clones contained microsatellites after the enrichment. In the present experiment, 100% of the sequenced positive clones possessed microsatellites. However, only 18% of the microsatellites were tetranucleotide repeats for which the enrichment was originally performed. The majority of the clones (about 78.63%) carried dinucleotide repeats,

particularly GT which is known to be the most prevalent type of dinucleotide motif among animals (Hancock 1999). Similar results were also obtained by Naish *et al.* (1998) who observed predominantly dinucleotide repeats when enrichment was carried out with a tetranucleotide GATA motif. Out of 36 polymorphic loci in present experiment only one contained the tetranucleotide GTCT repeat. Even though the targeted tetranucleotide repeats appeared in 18% of the positive clones, many of them could not be used due to lack of sufficient flanking region or due to very low number of repeat units. As a consequence, the enrichment method used in the present case did not serve the intended purpose.

More than 50% of the microsatellite repeats in rohu were perfect, about 41% were imperfect and the rest (about 9%) were compound in nature. In common carp 66.7% of the clones contained perfect repeats, 15.5% imperfect repeats and 17.9% of the clones contained compound repeats (Crooijmans *et al.* 1997). In silver barb 73.2% of the repeats were perfect in nature, 9.8% were imperfect and 17% were compound repeats (Kamonrat *et al.* 2002). Compared to these other cyprinid species, in rohu the occurrence of perfect repeats was much lower while the presence of imperfect repeats was substantially higher. A similar distribution of the type of repeats, however, was observed in rainbow trout with 56.9% perfect, 31.4% imperfect and 11.7% compound repeats (reviewed by Crooijmans *et al.* 1997).

The number of repeats per microsatellite locus in rohu varied from 3-49. Only those with more than five repeats were used for primer design. The largest proportion of the microsatellite arrays (27.35%) showed a repeat number in the range of 11-15, followed by 6-10 (23.93%). Similar observations were found in common carp where 35% of loci

had a repeat number between 6 and 11 (Crooijmans *et al.* 1997). In silver barb the most common size of dinucleotide motif was 12 (Kamonrat *et al.* 2002).

The polymorphism of the microsatellite loci in present study was tested on a panel of 8-16 unrelated individuals. This sample was a mixture of fishes collected from three wild populations of rohu collected from rivers in Bangladesh. The broad origin of the sample increased the chance of detecting polymorphism of the loci. However, it is possible that if the markers are tested on a larger sample size, more loci would turn out to be polymorphic. Due to lack of time the markers could not be checked for Mendelian inheritance.

Before this work was started, no microsatellites had been reported for rohu. During the work, 12 microsatellites were reported for rohu by Das *et al.* (2005). The 36 polymorphic microsatellites originating from the present work along with the previously reported loci can boost research in the areas of population and conservation genetics and can substantially aid in the management of aquaculture populations. The large number of microsatellites isolated in the present work can also be used in generating polymorphic loci in other closely related species through cross amplification. This has already been attempted by one of the MSc students (Sanjeev John) of the Institute of Aquaculture. He attempted to cross amplify a selection of 15 of these loci on two other commercially important cyprinid species, namely *L. fimbriatus* and *T. khudree* and was able to generate eight and six polymorphic loci respectively for the two species. The result of the cross-amplification is summarized in Table 6.6.

Table 6.6: Summary of results of cross-species amplification of 15 rohu microsatellite markers on two other cyprinid species. All amplification was carried out at 2.0 mM $MgCl_2$ concentration and 200 µM of each dNTP per reaction unless otherwise indicated (John 2005).

	Labeo fimbriatus			*Tor khudree*		
Locus	T_a (^oC)	Allele no.	Allele size range (bp)	T_a (^oC)	Allele no.	Allele size range (bp)
*Lro*01	NA	–	–	NA	–	–
*Lro*02	NA	–	–	52-60	2	165-171
*Lro*03	49-55	2	231-235	50-60	7	218-264
*Lro*04	55-60*	9	72-277	NA	–	–
*Lro*05	50-63	8	188-214	NA	–	–
*Lro*08	NA	–	–	48	5	157-173
*Lro*10	50-63	3	148-157	NA	–	–
*Lro*11	55-58	6	179-192	NA	–	–
*Lro*12	NA	–	–	NA	–	–
*Lro*14	NA	–	–	NA	–	–
*Lro*25	52-60	6	209-221	50-60	4	130-179
*Lro*26	55-60	4	312-319	50-60	3	296-304
*Lro*31	50-60**	2	189-195	50-60	4	165-231
*Lro*33	NA	–	–	NA	–	–
*Lro*34	NA	–	–	–	–	–

NA means no amplification
* amplified with 2.0 mM $MgCl_2$ and 230 µM of each dNTP
** amplified with 1.5 mM $MgCl_2$ and 200 µM of each dNTP

CHAPTER 7

GENERAL DISCUSSION

7.1 Introduction

Carps contribute the major share (41.9%) to world aquaculture production (FAO, 2004). Asia is the major producer of carps. Carps also constitute the most important group of fish in freshwater capture fisheries in Asian countries. Apart from the importance of carps as good sources of animal protein, their farming and fisheries contributes to the livelihoods of many poor people in the developing countries of Asia.

Overexploitation of fish from nature and degradation of natural habitats have depauperated many wild stocks. The burden of meeting the increasing demand for fish has therefore, shifted onto aquaculture production. As a result of this, carp aquaculture has flourished rapidly but several aspects of this growth have not occurred in a planned and scientific manner. Most of the carp culture farms and hatcheries are operated without following any scientific management protocols. The impacts of improper management have been detailed in Chapter 1. The present thesis focused on different genetic management aspects in carp aquaculture by using microsatellite markers. The work was carried out on three carp species, namely, silver carp, common carp and rohu. This chapter discusses the following topics:

- The achievements of the present research work and their implications.
- Future work and measures that may be taken to ensure conservation of wild populations and management and improvement of cultured stocks of carps.
- Breeding plan for setting up a selection programme

7.1.1 Achievements of present PhD work and future implications

7.1.1.1 Silver carp

- A number of newly isolated silver carp microsatellite markers were characterized. Two pentaplex PCR reactions using 10 of these loci were designed and optimized and were used for rapid genotyping of large numbers of silver carp individuals.
- The utility of these markers in sib-ship and relatedness analyses were assessed by applying them to a number of groups of fishes with known relationship categories. The efficiency of the markers in pedigree analysis was found to be good to moderate depending on the type of relationship categories involved. It was recommended that for differentiating between close relationship categories, such as half-sib from full-sib, more markers should be incorporated.
- Based on the results of pedigree analysis, these microsatellites were applied to analyse the pairwise relationships in a group of silver carp breeders, which supposedly originated from a well managed population. The analysis allowed the assessment of the strength of the management protocol followed for maintaining the broodstock.
- Based on the present work it will now be possible to use these silver carp markers in hatcheries for preventing mating between close relatives. It will also be possible to bring into management those hatchery populations where no pedigree record was previously maintained by applying relatedness analysis.
- Ten microsatellites were successfully applied to assign parentage to full-sib and half-sib families reared communally from a very early stage. Successful single family assignment was achieved in more than 96% of the individuals.
- Growth data from silver carp half-sib and full-sib families (after determining family identity with microsatellites) were successfully used to estimate the

heritability of harvest weight and harvest length in silver carp. Communal rearing helped to generate more accurate estimates of heritability by removing variation due to environmental differences.

- Heritability estimates were found to be quite high both for harvest weight (0.65 ± 0.14) and harvest length (0.50 ± 0.13). These values indicate the possibility of creating faster growing strains of silver carp through directional selection. This information, therefore, could be used to plan a selection programme for this species in Bangladesh.
- Use of silver carp microsatellites allowed assessment of the efficiency of pairwise mating between silver carp breeders in maintaining a large effective population size.
- Preliminary work was performed on cross-amplification of ten silver carp microsatellites on ten other cyprinid species. Out of 100 marker-species combinations, successful amplifications were observed in 87 cases. If the amplified loci are confirmed to be polymorphic on other species they can be useful for genetic management of those species. Finding polymorphic loci through cross-amplification will be especially useful for species for which no microsatellites have, so far, been reported, such as mrigal, *L. fimbriatus*, bighead carp, black carp, grass carp and *T. khudree*.

7.1.1.2 Common carp

- With the use of only five microsatellite markers it was possible to monitor a common carp mass selection programme and track important changes in the population structure at different stages of the breeding and management process.
- The use of the microsatellites helped detect a large perturbation in the contribution of individual stocks at different stages of the selection process compared to that in the base population.
- Microsatellite based parentage analysis in the selected population allowed the contribution of different families and its effect on effective population size (N_e) and on rate of inbreeding to be assessed.
- Microsatellite-based monitoring allowed informed recommendations for increasing the N_e and correcting the poor representation of one major stock (Amur) in the selected population.
- Family identity of individual selected fish generated through the present work can now being utilized in avoiding breeding between siblings for production of the F_2 selected population.
- Allelic and genotypic profiles were used to monitor the loss of genetic variation as a combined result of random drift and selection. Changes in allele number and allelic distribution were observed after just one generation of this selection programme.

7.1.1.3 Rohu

- Primer pairs were designed for 52 microsatellite loci, isolated from a partial genomic library of rohu and the loci were characterized. Thirty six loci were found to be polymorphic when tested on 8-16 individuals.
- These newly generated microsatellite markers will be useful in future works regarding genetic management and conservation of rohu populations.
- Cross-species amplification of these microsatellites can be used to develop markers for other related species. In a cross-amplification attempt with 15 of these markers, 8 loci were found to be polymorphic in *L. fimbriatus* and 6 loci in *T. khudree* (John 2005). All 52 loci should be tested for cross-amplification to develop larger numbers of polymorphic markers in these two species or in other cyprinid species.

7.1.2 Direction of future work on carps in South Asia

In order to protect the huge natural resource of carps in South Asia and to achieve sustainability and steady growth in carp aquaculture practice, attention needs to be directed in the following three major areas:

- Analysis and conservation of natural resources of carps
- Genetic management of broodstocks in hatcheries
- Genetic improvement of aquaculture stocks

These areas, however, are not mutually exclusive (described in Figure 7.1). Indeed the concordant progress in all the three areas is necessary to ensure the best utilization of available resources. The scope of future work in these areas is highlighted in the following sections with special emphasis on the potential uses of molecular markers.

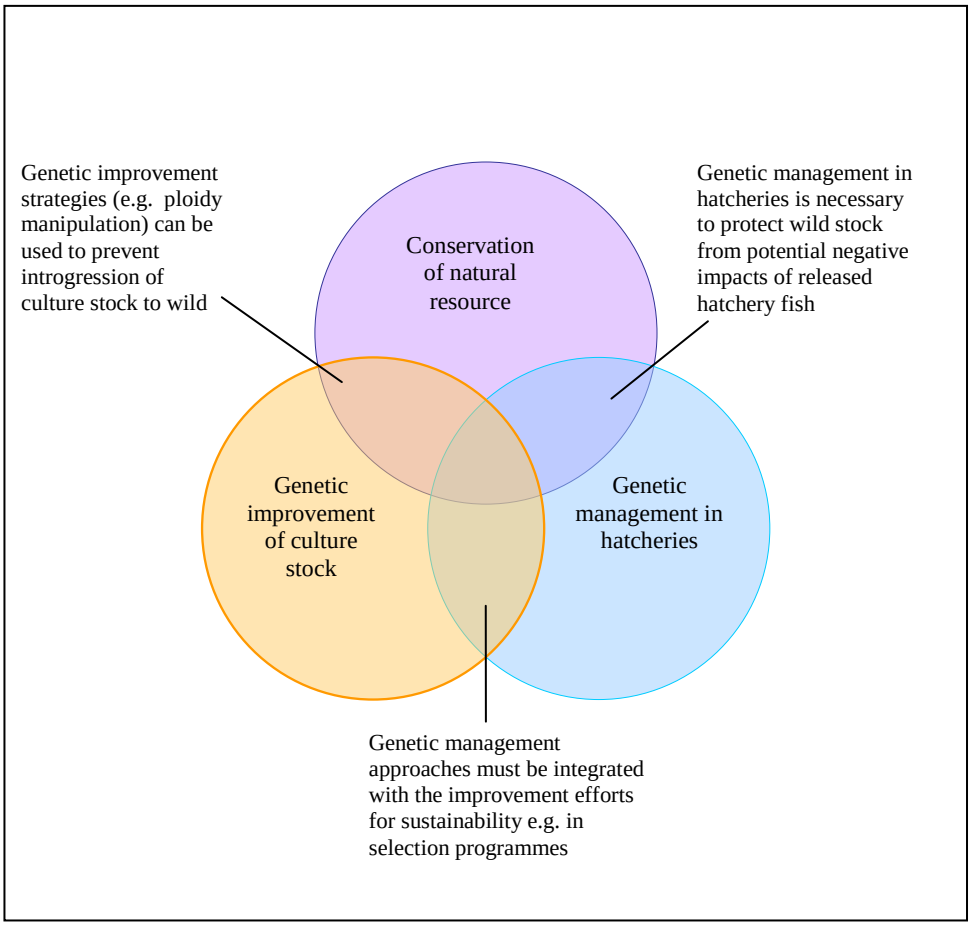

Figure 7.1: Three major areas of future work with examples of their possible overlap

7.1.2.1 Conservation of natural resources of carps

Many of the valuable aquatic species are now either extinct, endangered or are threatened in the wild. The World Conservation Union (IUCN) has listed 56 cyprinid species as “threatened” (From Acosta and Gupta 2005). In Bangladesh alone, 11 species of carps have been declared to be endangered while the natural stocks of many other species are declining (Hussain and Mazid 2001). A large number of factors have interacted to jeopardise the existence of these species, including destruction of natural breeding and feeding grounds of fishes through human intervention, blocking of migratory routes, pollution of water bodies and overfishing (Acosta and Gupta 2005). In order to set up effective conservation policies for protection of the valuable natural

resources of carps, information is required on the genetic structures of wild populations (Moran 2002, Primmer 2005). Another crucial area of research is the impact of hatchery fish on wild populations when they gain entry to nature either by escapes from farms or by governmental fry release programmes. Molecular markers are invaluable tools for research in all these areas.

Population genetic structure

Most species consist of a number of populations or stocks (Carvalho and Hauser 1995), which gradually become genetically distinct through their geographic, temporal or behavioural isolation. The lack of gene flow promotes local adaptation (Balloux and Lugon-Moulin 2002) through development of co-adaptive gene (allele) complexes. Conservation of this intraspecific diversity is important to preserve the flexibility of a species for adaptation to changes in the environment (Carvalho and Hauser 1995). It is, therefore, important to understand the population genetic structure of a species and to develop proper management policies to protect them from overexploitation or depletion through other causes.

Molecular markers have been successfully applied for detection, conservation and management of Pacific salmon stocks and species – Chinook salmon, Pink salmon (*Oncorhynchus gorbuscha*) and Chum salmon (*Oncorhynchus keta*) (Shaklee *et al.* 1999). MtDNA variation was used to identify the populations of Chinese carps viz. silver carp, bighead carp, grass carp and black carp in the major river systems of China and to formulate conservation policies (reviewed by Penman 2005). Similar research on carps in South Asia is scarce. Only recently, molecular markers have been applied to conduct some preliminary studies to distinguish among wild stocks of species. For instance RAPD markers have been used on rohu (*Labeo rohita*) samples from three

river systems in Bangladesh which have identified at least two distinct groups. Polymorphic microsatellites developed from cross-species amplification have been used to study population differentiation on mrigal (Lal *et al.* 2004) and *T. putitora* (Mohindra *et al.* 2004) in India. Deviation from Hardy-Weinberg equilibrium, variations in allele frequency and/or genotypic frequencies observed in both these studies indicated possible population differentiation. For *T. putitora* a small effective population size was also suspected. These studies, however, were only of preliminary nature and used small sample sizes. More systematic studies with larger sample sizes are required for conclusive assessment of population structure of these species. Furthermore, population genetics studies should be effectively integrated with the ecological and behavioural studies of the population as the success of nearly every conservation programme depends primarily on ecology and habitat (Moran 2002).

Another important aspect of population study is the estimation of effective population size (N_e) of individual stocks as this is an important index of genetic diversity and has important bearing on conservation decisions. Generally, the N_e is much smaller than the census population size (N). When information about number of breeders, sex ratios and variation in family size are known N_e can be estimated by direct methods. However, collecting such data from wild populations can be very difficult, if not impossible. A number of indirect methods based on molecular markers have, therefore, been developed which use temporal change in allele frequencies (Nei and Tajima 1981; Williamson and Slatkin 1999; Berthier *et al.* 2002), the excess of heterozygotes in progeny (Pudovkin *et al.* 1996; Luikart and Cornuet 1999) or linkage disequilibrium (Hill 1981; Bartely *et al.* 1992) for N_e estimation. So far there have not been any studies on effective population size of natural stocks of carps.

Enhancement of natural stocks

In order to enhance the capture production a large number of hatchery produced seed are released to open water bodies every year in India and Bangladesh. Although there has been widespread concern about the potential negative impact of poor quality hatchery fish on wild stock, so far no attempt has been made to study the impacts on carp populations. Exotic species can have even greater consequences as they may compete with native species and may disturb the ecosystem in natural water bodies (Bentsen and Thodesen 2005). This area, therefore, calls for urgent attention for research and monitoring. Molecular markers can be extremely useful in studying the introgression of the gene pool of hatchery fish into wild populations and in monitoring the changes in genetic constitution as a result of this. Research needs to be directed to find mechanisms to reduce the risk of negative impact of hatchery fish on wild population. Use of sterile fish such as triploidy has been suggested as an option to avoid introgression of stocked fish with wild population (Skibinski 1998). However, triploid fish can still cause considerable problems to local populations, for example through competition for food or if sterile males complete in courtship with fertile wild males. Besides, creation of triploid population may not be practical for certain species due to the nature of their breeding behaviour.

7.1.2.2 Genetic management of hatchery stocks

A major issue in the future of carp aquaculture in South Asia is how well we will be able to genetically manage the hatchery populations. Genetic management of hatchery stocks is important not only for the sustainability of aquaculture itself but also for the conservation of natural stocks, as large number of hatchery produced seeds mixes with wild population through release or escapes. The genetic quality of most captive hatchery carp broodstocks is thought to have eroded due to poor management as a result of lack

of awareness among the hatchery operators. Adopting corrective measures is the first step towards improvement of this situation. It is essential to educate the hatchery managers about the importance of genetic management and give them necessary training about good management practice. Besides, regulations need to be set to control the quality of seed produced in hatcheries and also to control the dissemination of the seeds.

Molecular markers will be useful under a number of circumstances especially for bringing the hatchery stock of exotic carps under management. While genetically poor broodstocks of native fish can still be replaced by new broodfish collected from wild, for exotic species this is almost impossible (Penman *et al.* 2002). In such cases, the possibility of applying the following mitigating options can be considered. Firstly, molecular markers, such as microsatellites can be used to analyse the kinship/relatedness between individuals of a broodstock in a hatchery. The marker identified relationship between individuals can then be recorded and maintained using PIT tags. Following this approach, it would be possible to prevent mating between closely related individuals and reduce inbreeding level. Secondly, molecular markers can be used evaluate the genetic relatedness/distance between broodstocks of different hatcheries and then crossbreeding between distant stocks can be performed to reduce the level of inbreeding (although there would be a chance of outbreeding depression). Thirdly, molecular markers can also be used to detect the presence of hybrid fish and to avoid their use as broodfish.

7.1.2.3 Genetic improvement of aquaculture stocks

Genetic improvement efforts with most of the commercially important carp species are only at their early stages. The majority of the research has been restricted to the areas of

interspecific or intergeneric hybridization, crossbreeding and chromosome manipulation (reviewed by Penman 2005). Although so far selective breeding has been applied only in a few cases for genetic improvement in carps, in future this will be the major avenue to exploit for creating improved aquaculture strains. Nevertheless, integration of other methods, e.g. hybridization, crossbreeding and genetic manipulation into selection programmes may boost the genetic improvement process (Dunham *et al.* 2000).

Most of the selection programmes on carps in South Asia have been performed on growth traits as currently these appear to be the most important traits from a commercial point of view. While growth related traits will definitely remain as the major phenotypes to be selected for, attempts should be made to incorporate other important traits as well. Recently in an attempt to identify the constraints upon increased yields in carp farming in Asia, Dey *et al.* (2005a) analysed the yield gap (the gap between observed and expected yields due to technological and/or socioeconomic constraints) and found that poor survival of carps in low oxygenated water and disease are the two major factors of low yield. Although improved farm management can reduce these problems considerably, these are difficult to overcome when the intensity level of farming is increased. Their study suggested that improved strains of carps should be produced which have higher disease resistance and can survive in low oxygenated water. Therefore, selection of these features along with growth traits can be considered. Studies need to be directed towards understanding genetic correlations between commercially important traits to evaluate the possibility of their simultaneous selection. Estimation of heritability of traits of different populations is another important area of study. Studies are also required on possible genotype-environment interaction so that strains with improved growth and survival in a broader range of environments can be created through a selection programme.

Although it still has not been applied to fish, marker-assisted selection (MAS) is opening a promising new line of research. The greatest promise of marker assisted selection is that it should increase the rates of genetic gain dramatically by increasing the accuracy of selection (Meuwissen 2003), although it is still to be proven by practical application. MAS should also be important in selecting traits with low heritability or traits which are difficult, expensive or impossible to record in a normal breeding programme. However, at present the major limitation of applying MAS is the number of genetic markers and QTL maps (Sonesson 2003).

7.1.3 Breeding plan for selection programme

Gjerde (2005) defines an optimally designed selection programme as one that maximises the genetic gain for a given trait over a certain period of time for a set of predefined constraints. Important constraints are the number of families and the size of the population that can be maintained in available facilities and the tolerable rate of inbreeding. Maintaining genetic diversity and keeping the inbreeding under control will be the major challenges in carp selection programmes. Careful breeding plans and selection designs will therefore, be required to cater with those needs. While designing a selective breeding programme for carp the following aspects should be considered.

7.1.3.1 Base population

A selection programme should be started from a population with broad genetic base. To create a base population individuals from several strains/stocks can be mixed or crossbred. Low selection intensity can be applied in the first generation of selection. Securing a broad genetic variability would allow long term selection response and a stepwise inclusion of new traits in the selection programme (Gjerde 2005).

7.1.3.2 Restricting inbreeding and maintaining genetic diversity

Accumulation of inbreeding is a major cause of reduction in genetic variance in a selected population (Gjerde *et al.* 1996). Methods to restrict the rate of inbreeding to an acceptable level include maintaining a large number of breeding individuals, homogenizing the sex ratio of the breeders and homogenizing the representation of families in selected population. According to Bentsen and Olesen (2002) to keep the inbreeding rates low (about 1% per generation) in a mass selection programme a minimum of 50 pairs of breeders (N_e =100) should be selected. Maintaining a uniform size of the families is a major challenge, especially in mass selection programmes. Although variability in family size probably cannot be circumvented totally because of highly variable fertilization and survival rates among families, the situation can be ameliorated through mixing of equal number of fry from each family before stocking them in communal rearing units. This strategy was used in the breeding programmes of silver carp and common carp (described in Chapters 4 and 5 respectively of the present thesis) and was found to be quite effective in maintaining a large N_e, compared to other breeding programmes where no such measures were taken. Apart from these strategies, molecular marker-based pedigree analysis can be used, wherever feasible, to reduce the variance in family representation and to avoid mating between relatives.

A compromise in selection intensity can also be considered to reduce the rate of inbreeding. High selection intensity results in rapid gain in response but that probably will not last for many generations. To ensure a long term response, selection intensity should be reduced. According to Falconer and Mackay (1996) selection of 50% would give the best compromise between selection gain and preservation of genetic diversity. Finally, maintaining independent selected lines and mating between them to produce

fish for commercial production can be applied to avoid the negative impact of inbreeding.

7.1.3.3 Mating design

Mating design is a crucial component of a breeding programme and its choice is eventually directed by the level of available facilities. Single pair mating is the easiest mating design and is useful in maintaining a 1:1 sex ratio, which is desirable in controlling the rapid reduction in $N_{e.}$ However, since only full-sibs are produced in this design, it is not possible to separate the additive genetic effect from non-additive effects (e.g. dominance, maternal and common environmental effects). This may result in biased breeding values with low accuracy, resulting in low genetic gain (Gjerde 2005). However, if the dominance and the maternal effects are low and if families can be reared in common ponds from an early stage, single pair mating can be an efficient means of fighting inbreeding. Factorial designs can overcome the limitations of pairwise mating but the serious drawback is that a lower number of males and females can be used for a given facility. Partial factorial designs may be a good alternative to full factorial designs. This design was applied for silver carp in the present PhD work and was found to be very promising. The design was technically easy to manage, allowed evaluation of both additive and non-additive effects and also allowed the use of reasonably large number of parents for a small sized hatchery.

7.1.3.4 Use of molecular markers in a breeding programme

The present PhD work has highlighted the advantage of using microsatellite markers in several aspects of a breeding programme. Firstly, the ability to assign parentage retrospectively using microsatellite markers allows rearing the families in communal ponds from a very early stage thus reducing the expense for maintaining families in

separate tanks and also preventing common environmental effects in creating bias in assessing the breeding values of individuals. The number of individuals and families can thus be increased without additional investment in creating new units for separate rearing. Secondly, microsatellite based kinship analysis can be used in homogenizing family representation in subsequent generations to maintain a high level of N_e and in controlling rate of inbreeding. Thirdly, changes in genetic diversity can be monitored using molecular markers. However, the use of molecular markers has some potential drawbacks. First of all, the technology of genotyping individuals at microsatellite loci is quite expensive. Secondly, low and variable survival rates among the families from the time of pooling the families to the time of selection may hugely increase the number of individuals to be typed to detect sufficient number of individuals from every family for selection (Gjerde 2005).

Therefore, whether molecular markers should be used in a breeding programme will depend on the specific needs and objectives of the programme and also on the availability of the resources. The combined use of DNA markers and physical tags can be applied for certain purposes such as for maintaining pedigree records of the individuals in a broodstock using PIT tags after the family identity of the individuals have been detected using molecular markers.

In using molecular markers costs are involved at two stages: the initial expense of setting up the facilities, which is an one-off expense and then the cost of analysing the samples. For hatcheries/farms in the developing countries of South Asia these expenses can become just enormous. However, the feasibility of building the facilities at certain key institutes can be considered, which can analyse the samples from different farms at reasonable charges.

7.2 Conclusion

The present thesis discusses genetic management practices in carp aquaculture and explores the applicability of molecular markers in breeding programmes and in hatchery population management. Since the breeding programmes used in the current PhD work were set in Bangladesh and India, they suit the context of the available resources in hatcheries of a developing country. It is true that the use of expensive molecular markers like microsatellites will not be a reality in most cases in the immediate future because of a lack of laboratory facilities and the cost involved. However, considering the benefits that the use of molecular markers can bring, attempts should be taken to develop the facilities at some key research institutes. Nonetheless, just by following certain rules for good hatchery practice such as proper designing of the breeding plan, maintaining a large population size with equal (or as close as possible) representation of both the sexes, avoiding the use of breeders from overlapping generations and avoiding inadvertent selection a great deal of improvement in genetic management can be ensured. Even if advanced molecular techniques are not used directly in such breeding schemes, the lessons learned from case studies involving molecular genetics can be applied.

REFERENCES CITED

Acosta, B.O. and Gupta, M.V. (2005) Threatened cyprinid species in Asia. pp. 138-145. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Ahmad, I., Bland, S.J.R., Price, C.R. and Kershaw, R. (1998) Open water stocking in Bangladesh: experiences from the Third Fisheries Project. *In: Inland Fisheries Enhancement: FAO Fisheries Technical paper number 374*, 463p.

Alam, M.S., Akanda, M.S.H., Khan, M.M.R. and Alam, M.S. (2002) Comparison of genetic variability between a hatchery and a river population of rohu (*Labeo rohita*) by allozyme electrophoresis. *Pakistan Journal of Biological Science* 5, 959-961.

Alarcòn, J.A., Magoulas, A., Georgakopoulos, T., Zouros, E., Alvarez, M.C. (2004). Genetic comparison of wild and cultivated European populations of the gilthead sea bream (*Sparus aurata*). *Aquaculture* 230, 65-80.

Allendorf, F.W. and Ryman, N. (1987) Genetic management of hatchery stocks. pp. 141-59. In: Ryman, N. and Utter, F., (eds.) *Population Genetics and Fishery Management*, Seattle: University of Washington press.

Aparicio, S. Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. Gelpke, M.D., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, B. and Brenner, S. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes. Science* 297, 1301-1310.

Armour, J.A., Neumann, R., Gobert, S. and Jeffreys, A.J. (1994) Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics* 3, 136-144.

Aulstad, D., Gjedrem, T. and Skjervold, H. (1972) Genetic and environmental sources of variation in length and weight of rainbow trout (*Salmo gairdneri*). *Journal of Fisheries Research Board of Canada* 29, 237-241.

Ayyappan, S., Ponniah, A.G., Reddy, P.V.G.K, Jana, R.K., Mahapatra, K.D. and Basavaraju, Y. (2001) Aquaculture genetic research in India: an overview. pp. 43-50. In: Gupta, M.V. and Acosta, B.O. (eds.) *Fish Genetics research in Member Countries and Institutions of the International Network on Genetics in Aquaculture. ICLARM conference proceedings* 64, 179p.

Balloux, F. and Lugon-Moulin, N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11, 155-165.

Banik, R.C.K. and Humayun, N. (1998) Fishery resources statistics. pp. 100-102. In: *Fish Week '98 Compendium.* Department of Fisheries and Ministry of Fisheries and Livestock, Dhaka.

Barman, H.K., Barat, A., Yadav, B.M., Banerjee, S., Meher, P.K., Reddy, P.V.G.K. and Jana, R.K. (2003) Genetic variation between four species of Indian major carps as revealed by random amplified polymorphic DNA assay. *Aquaculture* 217, 115-123.

Bártfai, R., Egedi, S., Yue, G.H., Kovács, B., Urbányi, B., Tamás, G., Horváth, L. and Orbán, L. (2003) Genetic analysis of two common carp broodstocks by RAPD and microsatellite markers. *Aquaculture* 219, 157-167.

Bartley, D., Bagley, M., Gall, G. and Bentley, B. (1992) Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. *Conservation Biology* 6, 365-375.

Basavaraju, Y., Penman, D.J. and Mair, G.C. (2002) Genetic status and strategies for improvement of common carp (*Cyprinus carpio*) in Karnataka, India-evaluation of stocks for the development of a breeding programme. pp. 29-36. In: Penman, D.J., Hussain, M.G., McAndrew, B.J. and Mazid, M.A., (Eds.) *Proceedings of a Workshop on Genetic Management and Improvement Strategies for Exotic Carps in Bangladesh.* 12-14 February 2002, Dhaka, Bangladesh: Bangladesh Fisheries Research Institute, Mymensingh.

Basavaraju, Y., Penman, D.J. and Mair, G.C. (2003) Stock evaluation and development of a breeding program for common carp (*Cyprinus carpio*) in Karnataka, India: progress of a research project. *NAGA WorldFish Center Quarterly* 26, 30-32.

Beacham, T.D. (1991) Developmental stability, heterozygosity, and genetic analysis of morphological variation in pink salmon (*Oncorhynchus gorbuscha*). *Canadian Journal of Zoology* 69, 274-278.

Beckman-Coulter (2004) Tailed primers methods for CEQ^{TM}2000 XL and 8000. Available at **www.mafes.msstate.edu/biotech/TailedPrimers.pdf**

Benson, G. (1999) Tandem repeat finder: a program to analyse DNA sequences. *Nucleic Acid Research* 27, 573-580.

Bentsen H.B. and Olesen, I. (2002) Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture* 204, 349-359.

Bentsen, H.B. and Thodesen, J. (2005) Genetic interactions between farmed and wild fish, with examples from the Atlantic salmon case in Norway. pp. 319-334. In: T. Gjedrem (ed.) *Selection and Breeding Programs in Aquaculture*. Springer, Netherlands.

Bentsen, H.H. and Gjerde, B. (1994) Design of fish breeding programs. pp. 353-359. In: *Proceedings of the 5th World Congress of Genetics Applied to Livestock Production.* Ontario: Department of Animal Science, University of Guelph.

Benzie, J.A.H., Kenway, M. and Trott, L. (1997) Estimates for the heritability of size in juvenile *Penaeus monodon* prawns from half-sib matings. *Aquaculture* 152, 49-53.

Berg, P. and Henryon, M. (1998) A comparison of mating designs for inference on genetic parameter in fish. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia, 11-16th January* 27,115-118.

Berthier, P. Beaumont, M.A., Cornuet, J.M. and Luikart, G. (2002) Likelihood-Based estimation of the effective population size using temporal changes in allele frequencies: A genealogical approach. *Genetics* 160, 741-751.

Bijma, P., Van Arendonk, J.A.M. and Woolliams, J.A. (2000) A general procedure for predicting rates of inbreeding in populations undergoing mass selection. *Genetics* 154: 1865-1877.

Birky, C.W., Fuerst, P., and Maruyama, T. (1989). Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effect of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121, 613-627.

Blanc, J.M. (2003) Comparison of experimental designs for estimating quantitative genetic parameters in fish. *Aquaculture Research* 34, 1099-1105.

Blouin, M.S. (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology and Evolution* 18, 503-511.

Blouin, M.S., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* 5: 393-401.

Borrell, Y.J., Álvarez, J., Vázquez, E., Pato, C.F., Tapia, C.M., Sánchez, J.A. and Blanco, G. (2004) Applying microsatellites to the management of farmed turbot stocks (*Scophthalmus maximus* L.) in hatcheries. *Aquaculutre* 241, 133-150.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32, 314-331.

Brody, J. R. and Kern, S. E. (2004) Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques* 36, 214-216.

Brown, C., Woolliams, J.A. and McAndrew, B.J. (2005) Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata. Aquaculture* 247, 219-225.

Brown, W.M. 1985. The mitochondrial genome of animals. pp. 95-130. In: MacIntyre, R.J. (ed.). *Molecular Evolutionary Genetics.* Plenum, NewYork.

Busack, C.A. and Currens K.P. (1995) Genetic risks and hazards in hatchery operations: fundamental concepts and issues. *American Fisheries Society Symposium* 15, 71-80.

Campton, D.E. (1995) Genetic effects of hatchery fish on wild populations of Pacific salmon and steelhead: what do we really know? *American Fisheries Society Symposium* 15, 337-353.

Carvalho, G.R. (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology* 43, 53-73.

Carvalho, G.R. and Hauser, L. (1995) Molecular genetics and the stock concept in fisheries. In: Carvalho, G.R. and Pitcher, T.J., (Eds.) *Molecular Genetics in Fisheries*, 1st edn. pp. 55-79. London: Chapman & Hall.

Carvalho, G.R. and Hauser, L. (1998) Advances in the molecular analysis of fish population structure. *Italian Journal of Zoology* 65, Supplement, 21-33.

Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Riaza, A., Ferreiro, I., Sánchez, L. and Martínez, P. (2004) Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture* 242, 119-135.

Charo-Karisa, H., Rezk, M.A., Bovenhuis, H. and Komen, H. (2005) Heritability of cold tolerance in Nile tilapia, *Oreochromis niloticus*, juveniles. *Aquaculture*, 249,115-123.

Chevassus, B., Krieg, F., Guyomard, R., Blanc, J.-M. and Quillet, E. (1992) The genetics of brown trout: twenty years of French research. *Icelandic Agricultural Science* 6, 109-124.

Chilcote, M.W., Leider, S.A. and Loch J.J. (1986) Differential reproductive success of hatchery and wild summer-run steelhead under natural conditions. *Transactions of the American Fisheries Society* 115, 726-735.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N. and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and Genbank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 100, 713-722.

Chow, S., Kishino, H. (1995) Phylogenetic relationships between tuna species of the genus *Thunnus* (*Scombriidae: Teleosrei*): inconsistent implications from morphology, nuclear and mitochondrial genome. *Journal of Molecular Evolution* 41, 741-748.

Clifford, S.L, McGinnity, P. and Ferguson, A. (1998) Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. *Journal of Fish Biology* 52, 118-127.

Cnaani, A., Hallerman, E.M., Ron, M., Weller, J.I., Indelman, M., Kashi, Y., Gall, G.A.E. and Hulata, G. (2003) Detection of chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size in an F_2 tilapia hybrid. *Aquaculture* 223, 117-128.

Crozier, W.W. (1998) Genetic implications of hatchery rearing in Atlantic salmon: effects of rearing environment on genetic composition. *Journal of Fish Biology* 52, 1014-1025.

Coughlan, J.P, Imsland, A.K., Galvin, P.T., Fitzgerald, R.D., Naevdal, G. and Cross, T.F. (1998). Microsatellite DNA variation in wild populations and farmed strains of turbot from Ireland and Norway: a preliminary study. *Journal of Fish Biology* 52, 916-922.

Crooijmans, R.P.M.A., Bierbooms, V.A.F., Komen, J., Vand Der Poel, J.J. and Groenen, M.A.M. (1997) Microsatellite markers in common carp (*Cyprinus carpio* L.). *Animal Genetics* 28, 129-134.

Cross, T.F. (1999) Genetic considerations in stock enhancement and sea ranching of invertebrates, marine fish and salmonids. pp. 37-48. In: Howell, B.R., Moksness, E. and Svasand, T., (eds.) *Stock Enhancement and Sea Ranching*. Oxford: Fishing News Books.

Cross, T.F. (2000) Genetic implications of transaction and stocking of fish species with particular reference to Western Australia. *Aquaculture Research* 31, 83-94.

Dakin, E.E. and Avise, J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93, 504-509.

Dan, N.C. Thien, T.M., Tuan and P.A. (2005) Carp genetic resources of Vietnam. pp. 72-81. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Danzmann, R.G. (1997) PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *Journal of Heredity* 88,333.

Das, P., Barat, A., Meher, P.K., Ray, P.P. and Majumdar, D. (2005) Isolation and Characterization of polymorphic microsatellites in *Labeo rohita* and their cross-species amplification in related species. *Molecular Ecology Notes* 5, 231-233.

DeWoody, J.A. and Avise J.C. (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared to other animals. *Journal of Fish Biology* 56, 46-473.

Dey, M.M., Alam, F. Miao, W., Piumsombun, S., Bhatta, R., Dung, L.T.C. and Paraguas, F.J. (2005a) Constraints to increased yields in carp farming: implications for future genetic research. pp. 114-120. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Dey, M.M., Paraguas, F.J., Bhatta, R., Alam, F., Weimin, M., Piumsombun, S., Koeshandrajana, S. Dung, L.T.C. and Sang, N.V. (2005b) Carp production in Asia: past trends and present status. pp. 6-15. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Doupé, R.G. and Lymbery, A.J. (2005) Additive genetic and other sources of variation in growth traits of juvenile black bream *Acathopagrus butcheri. Aquaculture research* 36, 621-626.

Duchesne, P., Godbout, M-H. and Bernatchez L. (2002) PAPA (Package for the Analysis of Parental Allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes* 2, 191-193.

Dunham, R.A. (2004) Aquaculture and Fisheries Biotechnology: Genetic approaches. CABI publishing, UK and USA.

Dunham, R.A., Majumdar, K., Hallerman, E., Bartley, D., Mair, G., Hulata, G., Liu, Z., Pongthana, N., Bakos, J., Penman, D., Gupta, M., Rothlisberg, P. and Hoerstgen-Schwark, G. (2001) Review of the status of aquaculture genetics. pp. 137-166. In: Subasinghe, R.P., Bueno, P., Phillips, M.J., Hough, C., McGladdery, S.E. and Arthur, J.R. (eds.) *Aquaculture in the Third Millennium. Technical Proceedings of the Conference on Aquaculture in the Third Millennium, Bankok, Thailand, 20-25 February 2000. NACA, Bangkok and FAO, Rome.*

Dupont-Nivet, M., Vandaputte, M. and Chevassus, B. (2002) Optimization of factorial mating designs for inference on heritability in fish species. *Aquaculture* 204, 361-370.

Ellegren, H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* 16, 551-558.

Elo, K. Ivanoff, S, Vuorinen, J.A., Piironen. J. (1997). Inheritance of RAPD markers and detection of interspecific hybridization with brown trout and Atlantic salmon. *Aquaculture* 152, 55-65.

Emlen, J.M. (1991) Heterosis and outbreeding depression: a multi-locus model and application to salmon production. *Fisheries Research* 12, 187-212.

Estoup, A., Jarne, P. and Cornuet, J.M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11, 1591-1604.

Evans, B., Bartlett, J., Sweijd, N., Cook, P. and Elliott, N.G. (2004). Loss of genetic variation at microsatellite loci in hatchery produced abalone in Australia (*Haliotis rubra*) and South Africa (*Haliotis midae*). *Aquaculture* 233, 109-127.

Ezenwa, V.O., Peters, J.M., Zhu, Y., Arévalo, E. Hastings, M.D., Seppa, P., Pedersen, J.S., Zacchi, F., Queller, D.C. and Stassmann, J.E. (1998) Ancient conservation of trinucleotide microsatellite loci in Polistine Wasps. *Molecular Phylogenetics and Evolution* 10, 168-177.

Falconer, D.S. and Mackay, F.C. (1996) Introduction to quantitative genetics. 4th edn. London: Prentice Hall.

FAO (2004) World Review of Fisheries and Aquaculture. Available at *http://www.fao.org///DOCREP/007/y5600e/y5600e04.htm*

FAO (2005) http://www.fao.org/fi/statist/statist.asp.

Ferguson, A. (1980) Biochemical systematics and evolution. 1st edn. Blackie & Son Limited, Glasgow.

Ferguson, M. (1995) The role of molecular genetic markers in the management of cultured fishes. In: Carvalho, G.R. and Pitcher, T.J., (Eds.) *Molecular Genetics in Fisheries*, 1st edn. pp. 81-103. London: Chapman & Hall.

Ferguson, M.M. and Danzmann, R.G. (1998) Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Sciences* 55, 1553-1563.

Fishback, A.G., Danzmann, R.G., Ferguson, M.M. and Gibson, J.P. (2002) Estimates of genetic parameters and genotype by environment interactions for growth traits of rainbow trout (*Oncorhynchus mykiss*) as inferred using molecular pedigree. *Aquaculture* 206, 137-150.

Fishback, A.G., Danzmann, R.G., Sakamoto, T. and Ferguson, M.M. (1999) Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 172, 247-254.

Fishbase (2005) http://www.fishbase.org

FitzSimmons, N.N., Moritz, C. and Moore, S.S. (1995) Conservation and Dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* 12, 432-440.

Fiumera, A.C., Parker, P.G. and Fuerst, P.A. (2000) Effective population size and maintenance of genetic diversity in captive-bred populations of a lake Victoria Cichlid. *Conservation Biology* 14, 886-892.

Fjalestad, K.T. (2005) Selection methods. pp. 159-173. In: T. Gjedrem (ed.) *Selection and Breeding Programs in Aquaculture*. Springer, Netherlands.

Gall, G.A.E. and Huang, N. (1988) Heritability and selection schemes for rainbow trout: body weight. *Aquaculture* 73, 43-56.

Garcia de Leon, F.J., Canonne, M., Quillet, E., Bonhomme, F. and Chatain, B. (1998) The application of microsatellite markers to breeding programmes in the sea bass, *Dicentrachus labrax*. *Aquaculture* 159, 303-316.

Gerber, S., Mariette, S., Streiff, R., Bodénès, C. and Kremer, A. (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* 9, 1037-1048.

GESAMP (1991) Reducing environmental impacts on coastal aquaculture. *Reports and Studies No. 47*, FAO, Rome 35p.

Gilbey, J., Verspoor, E. McLay, A. and Houlihan, D. (2004) A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Animal genetics* 35, 98-105.

Gjedrem, T. (1983) Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* 33, 51-72.

Gjedrem, T. and Olesen, I. (2005) Basic Statistical parameters. pp. 45-72. In: T.Gjedrem (ed.) *Selection and Breeding Programs in Aquaculture*. Springer, Netherlands.

Gjerde, B. (2005) Design of breeding programs. pp. 173-196. In: T. Gjedrem (ed.) *Selection and Breeding Programs in Aquaculture*. Springer, Netherlands, p 364.

Gjerde, B. and Gjedrem, T. (1984) Estimates of phenotypic and genetic parameters for carcass traits in Atlantic salmon and rainbow trout. *Aquaculture* 36, 97-110.

Gjerde, B. and Schaeffer, L.R. (1989) Body traits in rainbow trout. II. Estimates of heritabilities and of phenotypic and genetic correlations. *Aquaculture* 80, 25-44.

Gjerde, B., Gjøen, H.M. and Villanueva, B. (1996) Optimum designs of fish breeding programmes with constrained inbreeding. Mass selection for a normally distributed trait. *Livestock Production Science* 47, 59-72.

Goodnight K.F. and Queller, D.C. (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology* 8, 1231-1234.

Greelings, H., Van Oeveren, A.J., Pot, J.E. and Van Schaik, R.C. (1999) AFLP-Quantar™ Pro Image analysis software. *http: //www.keygene-products.com*.

Gunnes, K. and Gjedrem, T. (1981) A genetic analysis of body weight and length in rainbow trout reared in seawater for 18 months. *Aquaculture* 24, 161-174.

Hakanson, L., Ervik, A., Makinen, T. and Moller, B. (1988) Basic concepts concerning the environmental effects of marine fish farms. *Nord* 90, *Nordic Council of Ministers, Copenhagen*. 103p.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.

Hancock, J.M. (1995) The contribution of slippage-like processes to genome evolution. *Journal of Molecular Evolution* 41, 1038-47.

Hancock, J.M. (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. pp. 3-9. In: Goldstein, D.B. and Schlötterer, C., (Eds.) *Microsatellites: Evolution and Applications*, 1st edn. New York: Oxford University Press, Inc.

Hansen, M.M. (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite: an approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology* 11, 1003-1015.

Hansen, M.M., Loeschcke, V. (1996) Effects of releasing hatchery-reared brown trout to wild trout populations. *Biological Conservation* 75, 312

Hara, M. and Sekino, M. (2003) Efficient detection of parentage in a cultured Japanese flounder *Paralichthys olivaceus* using microsatellite DNA marker. *Aquaculture* 217, 107-114.

Hardy, O. and Vekemans, X. (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2, 618-620.

Hedrick, P.W., Brussard, P.F., Allendorf, F.W., Beardmore, J.A. and Orzack, Z. (1986) Protein variation, fitness and captive propagation. *Zoo Biology* 5, 91-99.

Henryon, M., Jojumsen, A., Berg, P., Lund, I., Pedersen, P.B., Olesen, N.J. and Slierendrecht, W.J. (2002) Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within farmed population of rainbow trout. *Aquaculture* 209, 59-76.

Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.T., Morris, D.B., Wright, J.M. and Cook, D. (1995) DNA fingerprint based analysis of parental and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245-256.

Herbinger, C.M., O'Reilly, P.T., Doyle, R.W., Wright, J.M. and O'Flynn, F. (1999) Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks. *Aquaculture* 173, 105-116.

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

Hindar, K., Ryman, N. and Utter, F. (1991) Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Sciences* 48, 945-957.

Hoffman J.I. and Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14, 599-612.

Hussain, M.G. and Mazid, M.A. (2001) Genetic improvement and conservation of carp species in Bangladesh. *Bangladesh Fisheries Research Institute and International Centre for Living Aquatic Resources Management*, 74p.

Hussain, M.G. and Mazid, M.A. (2005) Carp genetic resources of Bangladesh. pp. 16-25. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Islam, M.S. and Alam, M.S. (2004) Randomly amplified polymorphic DNA analysis of four different populations of the Indian major carp, *Labeo rohita* (Hamilton). *Journal of Applied Ichthyology* 20, 407-412.

Jarne, P. and Lagoda, P.J.L. (1996) Microsatellite, from molecules to populations and back. *Trends in Ecology and Evolution* 11, 424-429.

Jerry, D.R., Preston, N.P., Crocos, P.J., Keys, S., Meadows, J.R.S, Li, Y. (2004) Parentage determination of Kuruma shrimp *Penaeus* (*Marsupenaeus*) *japonicus* using microsatellite markers. *Aquaculture* 235, 237-247.

Jhingran, V.G. (1991) Fish and Fisheries of India. 3rd edn. Hindustan Publishing, Delhi.

Jiang, Z., Jianxin, W., Yongsheng, G. and Jiaxin, C. (2005) Carp genetic resources of China. pp. 26-38. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

John, S. (2005) Characterization of microsatellite loci in the Rohu, *Labeo rohita* and their cross-species amplification in other South Asian cyprinids. MSc thesis, submitted to the Institute of Aquaculture, University of Stirling.

Johnsen, B.O. and Jensen, A.J. (1994) The spread of furunculosis in Norwegian rivers. *Journal of Fish Biology* 45, 47-57.

Johnson, M.S. (2000) Measuring and interpreting genetic structure to minimise the genetic risks of translocations. *Aquaculture Research* 31, 133-143.

Jones A.G. (2001) GERUD 1.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays using multilocus DNA data. *Molecular Ecology Notes* 5, 708-711.

Jones, A.G. and Arden, W.R. (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* 12, 2511-2523.

Jordan, P.W. Goodman, A.E. and Donnellan, S. (2002) Microsatellite primers for Australian and New Guinean pythons isolated with an efficient marker development method for related species. *Molecular Ecology Notes* 2, 78-82.

Kamonrat, W., McConnell, K.J. and Cook, J. (2002) Polymorphic microsatellite loci from the Southeast Asian cyprinid, *Barbodes gonionotus* (Bleeker). *Molecular Ecology Notes* 2, 89-90.

Kandpal, R.P., Kandpal G. and Weissman, S.M. (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific marker. *Proceedings of the National Academy of Sciences of the USA* 91, 88-92.

Karagyozov, L., Kalcheva, I.D. and Chapman, V.M. (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acid Research*, 21, 3911-3912.

Kijas, J.M., Fowler, J.C., Garbett, C.A. and Thomas, M.R. (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligoneucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques* 16, 656-662.

Kocher, T.D., Albertson, R.C., Carleton, K.L. and Streelman, J.T. (2002) The genetic basis of biodiversity: genomic studies of cichlid fishes. pp. 35-44 In: Shimizu, N., Aoki, T., Hirono, I. and Takashima, F. (eds.) *Aquatic Genomics: Step Toward a Great Future*. Springer-Verlag, New York.

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D. and McAndrew, B. (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* 148, 1225-1232.

Koedprang, W., Ohara, K. and Taniguchi, N. (2000) Genetic and environmental variances on growth and reproductive traits of silver crucian carp *Carassius langsdorfii* using communal and separate rearing systems. *Fisheries Science* 66, 1092-1099.

Kohlmann, K., Gross, R., Murakaeva, A. and Kersten, P. (2003) Genetic variability and structure of common carp (*Cyprinus carpio*) populations throughout the distribution range inferred from allozyme, microsatellite and mitochondrial DNA markers. *Aquatic Living Resources* 16, 421-431.

Koljonen, M.L., Tähtinen, J., Marjatta, S. and Koskiniemi, J. (2002) Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture* 212, 69-92.

Konovalov, D.A., Manning C. and Heneshaw, M.T. (2004) KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes* 4, 779-782.

Kruuk, L.B. (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of Royal Society of London B.* 359, 873-890.

Lal, K.K., Chauhan, T., Mandal, A., Singh, R.K., Khulbe, L. Ponniah, A.G. and Mohindra, V. (2004) Identification of microsatellite DNA markers for population structure analysis in Indian major carp, *Cirrhinus mrigala* (Hamilton-Buchanan, 1882). *Journal of Applied Ichthyology* 20, 87-91.

Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* 43, 45-52.

Li, Y.C., Korol, A.B., Fahima, T., Beiles, A. and Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* 11, 2453-2465.

Liao, I.C. (1997) How can aquaculture help sustain world fisheries? p 431-436. In: Hancock, D.A., Smith, D.C., Grant, A. and Beumer, J.P., (Eds.) *2nd World Fisheries Congress Proceedings on Developing and Sustaining World Fisheries Resources: The State of Science and management, Australia.* CSIRO publishing.

Litt, M. and Luty J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44, 397-401.

Liu, X., Chang, Y., Xiang, J. and Cao, X. (2005) Estimation of genetic parameters for growth traits of the sea urchin, *Strongylocentrus intermedius. Aquaculture* 243, 27-32.

Liu, Z.J. and Cordes, J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1-37.

Liu, Z.J., Karsi, A., Li, P., Cao, D. and Dunham, R.A.(2003) An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* 165, 687-694.

Luikart, G. and Cornuet, J.M. (1999) Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* 151, 1211-1216.

Lynch, M and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753-1766.

Magoulas, A. (1998) Application of molecular markers to aquaculture and broodstock management with special emphasis on microsatellite DNA. pp. 153-168. In: Bartley, D. and Basurco, B., (eds.) *Genetics and breeding of Mediterranean aquaculture species: Proceedings of the seminar of the CIHEAM Network on Technology of Aquaculture in Mediterranean* (TECAM), Jointly organized by CIHEAM and FAO in Zaragoza (Spain) from 28-29 April 1997, CIHEAM.

Maguire, T.L. (2001) Producing and exploiting enriched microsatellite libraries. In: Henry, R.J. (ed.) *Plant Genotyping: the DNA Fingerprinting of Plants*, pp. 193-209. CAB International.

Mair, G.C., Penman, D.J., McAndrew, B.J., Hussain, M.G., Sattar, A., Basavaraju, Y. and Tuan, P.A. (2002) Genetic management and improvement strategies for exotic carps in Asia. *Progress report, Mid term review:* April 1, 2000 to October 31, 2002.

Marshall, T.C., Slate, J., Kruuk, L. and Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7**,** 639-655.

Masser, M.P. (2002) Using grass carp in aquaculture and private impoundments. *SRAC publication no. 3600. United States Department of Agriculture, Cooperative State Research, Education and Extension Service. Available at:* http://aquanic.org/publicat/usda_rac/efs/srac/3600fs.pdf

Matsuda, M. and Nagahama, Y. (2002) Positional cloning of the sex-determining region of medaka using a Y congenic strain. pp. 236-245. In: Shimizu, N., Aoki, T., Hirono, I. and Takashima, F. (eds.) *Aquatic Genomics: Step Toward a Great Future.* Springer-Verlag, New York.

McDonald, G.J., Danzmann, R.G., Ferguson, M.M. (2004) Relatedness determination in the absence of pedigree information in three cultured strains of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 233: 65-78.

McKay, L.R., Ihssen, P.E. and Friars, G.W. (1986) Genetic parameters of growth in rainbow trout, *Salmo gairdneri*, prior to maturation. *Canadian Journal of Genetics and Cytology* 28, 306-312.

Meuwissen, T. (2003) Genomic selection: the future of marker assisted selection and animal breeding. pp. 54-59. *Proceedings of the Workshop on Maker Assisted Selection: A fast Track to Increase Genetic Gain in Plant and Animal Breeding?* 17-18 October, 2003, Turin, Italy.

Meuwissen, T.H.E. and Woolliams, J.A. (1994) Effective sizes of livestock populations to prevent a decline in fitness. *Theoretical and Applied Genetics* 89, 1019-1026.

Meyer, K. (1989) Estimation of genetic parameters. pp. 161-167. In: W.G. Hill and T.F.M. McKay (eds), *Evolution and Animal Breeding: Reviews on Molecular and Quantitative Approaches in Honour of A. Robertson.*

Mia, M.Y, Taggart, J.B., Gilmour, A.E., Gheyas, A.A., Das, T.K., Kohinnor, A.H.M., Rahman, M.A., Sattar, M.A., Hussain, M.G., Mazid, M.A., Penman, D.J. and McAndrew, B.J. (2005). Detection of hybridization between Chinese carp species (*Hypophthalmichthys molitrix* and *Aristichthys nobilis*) in hatchery broodstock of Bangladesh, using DNA microsatellite loci. *Aquaculture* 247, 267-273.

Mignon-Grasteau, S., Boissy, A., Bouix, J., Faure, J.M., Fisher, A.D., Hinch, G.N., Jensen, P., Neindre, P.L., Mormède, Prunet, P., Vandaputte, M., Beaumont, C. (2005) Genetics of adaptation and domestication in livestock. *Livestock Production Science* 93, 3-14.

Milligan, B.G. (2003) Maximum-likelihood estimation of relatedness. *Genetics* 163, 1153-1167.

Moav, R. and Wohlfarth, G.W. (1976) Two way selection for growth rate in the common carp (*Cyprinus carpio* L.). *Genetics* 82, 83-101.

Mohindra, V., Ranjana, L., Khulbe, L., Ponniah, A.G. and Lal, K.K. (2004) Microsatellite loci to assess the genetic variation in *Tor putitora*. *Journal of Applied Ichthyology* 20, 466-469.

Mohindra, V., Anshumala, Punia, P., Narain, L., Kapoor, D. and Lal, K.K. (2005) Microsatellite loci to determine population structure of *Labeo dero* (Cyprinidae). *Aquatic Living Resources* 18, 83-85.

Moore, A.J. and Kukuk, P.F. (2002) Quantitative Genetic Analysis of Natural Populations. *Nature Reviews Genetics* 3, 971-978.

Moran, P. (2002) Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. *Ecology of Freshwater Fish* 11, 30-55.

Moritz, C. (1994) Defining evolutionary significant units for conservation. *Trends in Ecology and Evolution* 9, 373-375.

Mousseau, T.A., Ritland, K. and Heath, D.D. (1998) A novel method of estimating heritability using molecular markers. *Heredity* 80, 218-224.

Mridula, R.M., Manissery, J.M., Keshavanath, P., Shankar, K.M., Nandeesha, M.C. and Rajish, K.M. (2005) Effects of paddy straw and sugarcane bagasse on water quality, bacterial biofilm production, and growth and survival of rohu, *Labeo rohita* (Hamilton). *Aquaculture Research* 36, 635-642.

Naish, K.-A. and Skibinski, D.O.F.S. (1998) Tetranucleotide microsatellite loci from Indian carp. *Journal of Fish Biology* 53, 886-889.

Neff B.D., Repka J. and Gross M.R. (2001) A Bayesian framework for parentage analysis: the value of genetic and other biological data. *Theoretical Population Biology*, 59, 315-331.

Nei, M. and Tajima, F. (1981) Genetic drift and estimation of effective population size. *Genetics* 98, 625-640.

Nenashev, G.A. (1966) The determination of heritability of different characters in fishes. *Genetika* 11, 100-108.

Nichols, K.M., Young, W.P., Danzman, R.G., Robison, B.D. Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I, Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A. and Thorgaard, G.H. (2003) A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* 34, 102-115.

Nielsen, R., Mattila, D.K., Clapman P.J. and Palsbøll, P.J. (2001) Statistical approaches to paternity analysis in natural populations and applications to Northern Atlantic humpback whale. G*enetics* 157, 1673-1682.

Noakes, D.J., Beamish, R.J., Kent, M.L. (2000) On the decline of Pacific salmon and speculative links to salmon farming in British Columbia. *Aquaculture* 183, 363-386.

Norris A.T., Bradley, D.G. and Cunningham, E.P. (1999) Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture* 180, 247-264.

Norris, A.T., Bradley, D.G. and Cunningham, E.P. (2000) Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture* 182, 73-83.

O'Brien, S.J. (1991) Molecular genome mapping: lessons and prospects. *Current Opinion in Genetics & Development* 1, 105-111.

O'Connell, M. and Wright, J.M. (1997) Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* 7, 331-363.

O'Reilly, H.L.C., McConnell, K. and Wright, J.M. (1996) Rapid analysis of genetic variation in Atlantic salmon (S*almo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* 53, 2292-2298.

O'Reilly, P. and Wright, J.M. (1995) The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. *Journal of Fish Biology* 47, 29-55.

O'Reilly, P.T., Herbinger, C. and Wright, J.M. (1998) Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. *Animal Genetics* 29, 363-370.

Ostrander, E.A., Jong, P.M., Rine, J. and Duyk, G. (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequence. *Proceedings of the National Academy of Sciences of the USA* 89, 3419-3423.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M.R., Akutsu, T. and Okamoto, N. (2001) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics* 265, 23-31.

Padhi, B.K. and Mandal, R.K. (1994) Improper fish breeding practices and their impact on aquaculture and fish biodiversity. *Current Science* 66, 624-626.

Paetkau, D. (1999) Microsatellites obtained using strand extension: an enrichment protocol. *BioTechniques* 26, 690-697.

Painter, I. (1997) Sibship reconstructed without parental information. *Journal of Agricultural, Biological and Environmental Statistics* 2, 212-229.

Pante, M.J.R, Gjerde, B. and McMillan, I (2001) Inbreeding levels in selected populations of rainbow trout, *Oncorhynchus mykiss. Aquaculture* 192, 213-224.

Park, L.K. and Moran, P. (1995) Developments in molecular genetic techniques in fisheries. pp. 1-28. In: Carvalho, G.R. and Pitcher T.J., (eds.) Molecular Genetics in Fisheries, 1st edn. London: Chapman & Hall.

Penman, D.J. (2005) Progress in carp genetic research. pp. 82-113. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Penman, D.J. and McAndrew, B.J. (1998) Practical aspects of selection and fitness of stocked fish. *In: Inland Fisheries Enhancement: FAO Fisheries Technical paper number 374*, p463.

Penman, D.J., Mair, G.C., McAndrew, B.J., Hussain, M.G., Sattar, A., Basavaraju, Y. and Tuan, P.A. (2002) Genetic management and improvement strategies for exotic carps in Asia: a project overview. pp. 1-8. In: Penman, D.J., Hussain, M.G., McAndrew, B.J. and Mazid, M.A., (eds.) *Proceedings of a Workshop on Genetic Management and Improvement Strategies for Exotic Carps in Bangladesh.* 12-14 February 2002, Dhaka, Bangladesh: Bangladesh Fisheries Research Institute, Mymensingh.

Perez-Enriquez, R. Takagi, M. and Taniguchi, N. (1999) Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. *Aquaculture* 173, 413-423.

Perry, R.P., Danzmann, R.G., Ferguson, M.M. and Gibson, J.P. (2001) Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity* 86, 333-341.

Pongthana, N. (2005) Carp genetic resources of Thailand. pp. 63-71. In: Penman, D.J., Gupta, M.V. and Dey, M.M. (eds.) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Porta, J., Porta, M., Martínez-Rodriguez, G., and Alvarez, M.C. (2006) Genetic structure and genetic relatedness of a hatchery stock of Senegal sole (*Solea senegalensis*) inferred by microsatellites. *Aquaculture*, 251, 46-55.

Price, E.O. and King, J.A. (1968). Domestication and adaptation. pp. 34-45. In: Hafez, E.S.E. (ed.), *Adaptation of Domestic Animals.* Lea and Febiger, Philadelphia.

Primmer, C.R. (2005) Genetic characterization of populations and its use in conservation decision making in fish. pp. 137-142. *The Role of Biotechnology*. Villa Gualino, Turin, Italy-5-7 March, 2005.

Primmer, C.R. Møller, A.P. and Ellegen, H. (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* 5, 365-378.

Pudovkin, A.I., Zaykin, D.V. and Hedgecock, D. (1996) On the potential for estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* 144, 383-387.

Queller D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. *Evolution* 43, 258-275.

Rajts, F. Huntingdon, T. and Hussain, M.G. (2002) Carp broodstock management and genetic improvement programme under Fourth Fisheries Project. pp 95-106. In: Penman, D.J., Hussain, M.G., McAndrew, B.J. and Mazid, M.A., (Eds.) *Proceedings of a Workshop on Genetic Management and Improvement Strategies for Exotic Carps in Bangladesh.* 12-14 February 2002, Dhaka, Bangladesh: Bangladesh Fisheries Research Institute, Mymensingh.

Raymond M. and Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86, 248-249.

Reddy, P.V.G.K. (1999) Genetic resources of Indian major carps, their distribution and characterization. In: *Genetic Resources of Indian Major Carps: FAO Fisheries Technical paper number 387*, 76p.

Reddy, P.V.G.K. (2005) Carp genetic resources of India. pp. 39-53. . In: Penman, D.J., Gupta, M.V. and Dey, M.M. (Eds) *Carp Genetic Resources for Aquaculture in Asia*. *WorldFish Center Technical Report* 65, 152p.

Reed, D.H. and Frankham, R. (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55, 1095-1103.

Reed, D.H. and Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conservation Biology* 17, 230-237.

Rhymer, J.M. and Simberloff, D. (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27, 83-109.

Rico, C. Rico, I. and Hewitt, G. (1996) 470 million years of conservation of microsatellite loci among fish species. *Proceedings of Biological Science* 263, 549-557.

Ritland, K. (1996) Estimators of pairwise relatedness and inbreeding co-efficients. *Genetical Research* 67, 175-187.

Ritland, K. (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. *Molecular Ecology* 9, 1195-1204.

Rodriguez, M.F., LaPatra, S., Williams, S., Famula, T. and May, B. (2004) Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* 241, 93-115.

Rogne, S. (1995) Genetics and Breeding in Aquaculture. pp.117-124. In: Reinertsen, H. and Haaland, H., (eds.) *Proceedings of the First International Symosium on Sustainable Fish Farming.* Held in Oslo, Norway from 28-31 August, 1994. Netherlands, USA: A.A. Balkema/ Rotterdam/ Brookfield.

Rutten, M.J.M, Bovenhuis, H. and Komen, H. (2005) Genetic parameters for fillet traits and body measurements in Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* 246, 125-132.

Ruzzante, D.E. Taggart, C.E. and Cook, D. (1999) A review of the evidence of genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. *Fisheries Research* 43, 79-97.

Ryman N., Utter, F. and Hindar, k. (1995) Introgression, supportive breeding, and genetic conservation. pp. 341-365. In: Ballou, J.D., Gilpin, M. and Foose, T.J., (eds.) *Population Management for Survival and Recovery*, New York: Columbia University Press.

Sakamoto, T., Danzmann, R.G., Okamoto, N., Ferguson, M.M. and Ihssen, P.E. (1999) Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173, 33-43.

Sarkar, G., Paynton, C. and Sommer, S.S. (1991) Segments containing alternating purine and pyrimidine dinucleotides: patterns of polymorphism in humans and prevalence throughout phylogeny. *Nucleic Acid Research* 19, 631-636.

Sattar, M.A. and Das, T.K. (2002) Broodstock management of Chinese carps and dissemination strategy at NFEP, Parbatipur. pp. 43-50. In: Penman, D.J., Hussain, M.G., McAndrew, B.J. and Mazid, M.A., (eds.) *Proceedings of a Workshop on Genetic Management and Improvement Strategies for Exotic Carps in Bangladesh.* 12-14 February 2002, Dhaka, Bangladesh: Bangladesh Fisheries Research Institute, Mymensingh.

Sekino, M., Hara, M. and Taniguchi, N. (2002) Loss of microsatellite and mitochondrial DNA variation in hatchery strains of Japanese flounder *Paralichthys olivaceus*. *Aquaculture* 213, 101-122.

Sekino, M., Saitoh, K., Yamada, T., Kumagai, A., Hara, M. and Yamashita, Y. (2003) Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. *Aquaculture 221*, 255-263.

Sekino, M., Sugaya, T, Hara, M. and Taniguchi, N. (2004) Relatedness inferred from microsatellite genotypes as a tool for broodstock management of Japanese flounder *Paralichthys olivaceus. Aquaculture* 233, 163-172.

Shaklee, J.B., Beacham, T.D., Seeb, L. and White, B.A. (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research* 43, 45-78.

Shikano, T. (2005) Marker-based estimation of heritability for body color variation in Japanese founder *Paralichthys olivaceus. Aquaculture*, 249, 95-105.

Shikano, T. and Tabihuchi, N. (2002) Using microsatellite and RAPD markers to estimate the amount of heterosis in various strain combinations in the guppy (*Poecilia reticulata*) as a fish model. *Aquaculture* 204, 271-281.

Shimoda, N., Knapik, E.W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacobs, H. and Fishman, M.C. (1999) Zebrafish genetic map with 2000 microsatellite markers. *Genomics* 58, 219-232.

Signorovitch, J. and Nielsen, R. (2002) PATRI- paternity inference using genetic data. *Bioinformatics*, 18, 341-342.

Simonsen, V. Hansen, M.M., Sarder, M.R.I. and Alam, M.S.(2004) High level of hybridization in three species of Indian major carps. *NAGA, WorldFish Center Quarterly* 27, 65-69.

Skaala, Ø., Dahle, G., Jørstad, K.E. and Nævdal, G. (1990) Interactions between natural and farmed fish populations: Information from genetic markers. *Journal of Fish Biology* 36, 449-460.

Skibinski, D.O.F. (1998) Genetical aspects of fisheries enhancement. *In: Inland Fisheries Enhancement: FAO Fisheries Technical paper number 374*, p463.

Smith, B.R., Herbinger, C.M. and Merry, H.R. (2001) Accurate partitioning of individuals into full-sib families from genetic data without parental information. *Genetics* 158, 1329-1338.

Smith, P. and McVeagh, M. (2000) Allozyme and mirosatellite DNA markers of toothfish population structure in the Southern Ocean. *Journal of Fish Biology* 57 (supplement A), 72-83.

Sonesson, A.K. (2003) Possibilities for marker-assisted selection in fish breeding schemes. pp. 60-65. *Proceedings of the Workshop on Maker Assisted Selection: A fast Track to Increase Genetic Gain in Plant and Animal Breeding?* 17-18 October, 2003, Turin, Italy.

Stone J. and Björklund, M. (2001) DELIRIOUS: a computer program designed to analyze molecular marker data and calculate delta and relatedness estimates with confidence. *Molecular Ecology Notes* 1, 212-290.

Sun, X. and Liang, L. (2004) A genetic linkage map common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* 238, 165-172.

Takagi, M., Sato, J., Monbayashi, C., Aoki, K., Tsuji, T., Hatanaka, H., Takahashi, H. and Sakai, H. (2003) Evaluation of microsatellites identified in the tiger puffer *Takifugu rubripes* DNA database. *Fisheries Science* 69, 1085-1095.

Tanck, M.W.T., Palstra, A.P, Van De Weerd, M., Leffering, C.P., Van Der Poel, J.J., Bovenhuis, H. and Komen, J. (2001) Segregation of microsatellite alleles and residual heterozygosity at single loci in homozygous and androgenetic common carp (*Cyprinus carpio* L.) *Genome* 44, 743-751.

Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acid Research* 17, 6463-6471.

Tautz, D. and Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acid Research* 12, 4127-4138.

Tautz, D., Trick, M. and Dover, G.A. (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322, 652-656.

Tave, D. (1993) Genetics for fish hatchery managers, 2nd edn. New York: Van Nostrand Reinhold.

Taylor, E.B. (1991) A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* 98, 185-207.

Templeton, A.R. (1986) Coadaptation and outbreeding depression. pp. 105-116. In: Soulé, M.E., (ed.) *Conservation Biology: The Science of Scarcity and Diversity*. Washington DC: Sinauer Associates.

Tessier, N., Bernatchez, L. and Wright, J.M. (1997). Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analysis in land-locked Atlantic salmon *Salmo salar* L. *Molecular Ecology* 6, 735-750.

Thomas, S. and Hill, W.G. (2000) Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* 155, 1961-1972.

Thompson, E.A. (1975) The estimation of pairwise relationships. *Annals of Human Genetics* 39, 173-188.

Tong, J., Wang, Z., Yu, X., Wu, Q, and Chu, K.H. (2002) Cross-species amplification in silver carp and bighead carp with microsatellite primers of common carp. *Molecular Ecology Notes* 2, 245-247.

Tóth, G., Gáspári, Z. and Jurka, J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* 10, 967-981.

Van De Casteele, T., Galbusera, P. and Matthysen, E. (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology* 10, 1539-1549.

Vandeputte, M. (2003) Selective breeding of quantitative traits in the common carp (*Cyprinus carpio*): a review. *Aquatic Living Resources* 16, 399-407.

Vandeputte, M., Kocour, M, Mauger, S., Dupont-Nivet, M., de Guerry, D., Rodina, M., Gela, D., Vallod, D., Chevassus, B. and Linhart, O. (2004) Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture* 235, 223-236.

Vandeputte, M., Quillet, E., Chevassus, B. 2002. Early development and survival in brown trout (*Salmo trutta fario* L.): indirect effects of selection for growth rate and estimation of genetic parameters. *Aquaculture* 204, 435-445.

Villanueva, B., Verspoor, E. and Visscher, P.M. (2002) Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Animal Genetics* 33, 33-41.

Volckaert, F.A.M. and Hellemans, B. (1999) Survival, growth and selection in a communally reared multifactorial cross of African catfish (*Clarias gariepinus*). *Aquaculture 171*, 49-64.

Waldbieser, C.G., Bosoworth, Nonneman, D.J. and B.G., Wolters, W.R. (2001) A microsatellite-based genetic linkage map for channel catfish *Ictalurus punctatus. Genetics* 158, 727-734.

Wallin, J.M., Holt, C.L., Lazaruk, K.D., Nguyen, T.H. and Walsh, P.S. (2002) Constructing universal multiplex PCR systems for comparative genotyping. *Journal of Forensic Science* 47, 52-65.

Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. *Genetics* 160, 1203-1215.

Wang, Z., Weber, J.L., Zhong, G. and Tanksley, S.D. (1994) Survey of the plant short tandem DNA repeats. *Theoretical and Applied Genetics* 88, 1-6.

Waples, R.S. (1999) Dispelling some myths about hatcheries. *Fisheries* 24, 12-21.

Was, A. and Wenne, R. (2002) Genetic differentiation in hatchery and wild sea trout (*Salmo trutta*) in the Southern Baltic at microsatellite loci. *Aquaculture* 204, 493-506.

Weber, J.L. and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44, 388-96.

Williamson, E.G. and Slatkin, M. (1999) Using maximum likelihood to estimate size from temporal changes in allele frequencies. *Genetics* 152, 755-761.

Winkelman, A.M. and Peterson, R.G. (1994) Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in Chinook salmon. *Aquaculture* 125, 17-30.

Wright, J. (1993) DNA fingerprinting in fishes. pp. 57-91. In: Hochachka, P. and Mommsen, T., (eds.) *Biochemistry and Molecular Biology of Fishes*, vol. 2 Elsevier, Amsterdam.

Wright, J.M. and Bentzen, P. (1995) Microsatellites: genetic markers for the future. pp. 117-121. In: Carvalho, G.R. and Pitcher, T.J., (eds.) *Molecular Genetics in Fisheries*, 1st edn. London: Chapman & Hall.

Xu, Z., Primaver, J.H., de la Pena, L.D., Pettit, P., Belak, J. and Acivar-Warren, A. (2001) Genetic diversity of wild and cultured black tiger shrimp (*Penaeus monodon*) in the Philippines using microsatellites. *Aquaculture* 199, 13-40.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P. and Thorgaard, G.H. (1998) A detailed linkage map of rainbow trout produced using doubled haploids. *Abstracts in Plant and Animal Genome* 6, 165.

Zane, L., Bargelloni, L. and Patarnello, T. (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11, 1-16.

Zimmerman, A.M., Wheeler, P.A., Ristow, S.S. and Thorgaard, G.H. (2005) Composite interval mapping reveals three QTL associated with pyloric caeca number in rainbow trout, *Oncorhynchus mykiss. Aquaculture* 274, 85-95.

PAPERS FROM PRESENT THESIS

Mia, M.Y, Taggart, J.B., Gilmour, A.E., **Gheyas, A.A**., Das, T.K., Kohinnor, A.H.M., Rahman, M.A., Sattar, M.A., Hussain, M.G., Mazid, M.A., Penman, D.J. and McAndrew, B.J. (2005). Detection of hybridization between Chinese carp species (*Hypophthalmichthys molitrix* and *Aristichthys nobilis*) in hatchery broodstock of Bangladesh, using DNA microsatellite loci. *Aquaculture* 247, 267-273.

Gheyas, A.A., Cairney, M., Gilmour, A.E., Sattar, M.A., Das, T.K., McAndrew, B.J., Penman, D.J. and Taggart, J.B. (2006) Characterization of microsatelllite loci in silver carp (*Hypophthalmichthys molitrix*), and cross-amplification in other cyprinid species. *Molecular Ecology Notes* (Accepted).