University of Southampton

Molecular Markers and the Speciation of African Cichlid Fish

Charlotte Jane Allender

Submitted for the qualification of Doctor of Philosophy

Biodiversity and Ecology Division
School of Biological Sciences

October 2001

University of Southampton

Abstract

Faculty of Science

SCHOOL OF BIOLOGICAL SCIENCES

Doctor of Philosophy

MOLECULAR MARKERS AND THE SPECIATION OF AFRICAN CICHLID FISH

By Charlotte Jane Allender

The cichlid fish of the African Great Lakes represent an astounding example of explosive speciation and adaptive radiation. The rapid nature of the evolution of these species has provoked intense interest in the evolutionary mechanisms responsible for causing and maintaining divergence in this remarkable group of organisms.

The sexual dichromatism and breeding behaviour exhibited by some groups of African cichlids have led many researchers to infer that sexual selection through female preference for male colour patterns may be one such mechanism. Given the apparent importance of male colour patterns in mate choice, the specific status of allopatric populations that vary in male colour patterns is unclear. Microsatellite DNA was used to assess the paternity of offspring in two sets of mate choice trials where females chose between males that differ in colour pattern. One trial tested a sympatric species pair from Lake Victoria (*Pundamilia pundamilia* and *P. nyererei*), and the other tested two allopatric populations of the same putative species from Lake Malawi (*Metriaclima zebra*). In both cases, non-assortative mating was observed. However it was not clear if these results indicated genuine female choice or were a product of the design of the trials. Microsatellite DNA can also be used to generate statistical estimates of pairwise relatedness. The effect on such estimates of locus number and heterozygosity was assessed in two species of cichlids from Lake Malawi (*M. zebra* and *M. callainos*). A total of eleven loci were used in a combination of tests on known relatives and computer simulations. It was found that using a higher number of more variable loci resulted in more accurate estimates of relatedness. However it was observed that nine loci provided a similar degree of accuracy to that obtained using eleven loci. This result has obvious implications for the time and cost required for large-scale studies of relatedness in natural populations. Phylogenetic relationships between populations and species of rock-dwelling cichlids from Lake Malawi were reconstructed using both six microsatellite loci and AFLP markers. The microsatellite data set appeared to contain too few loci to resolve all relationships but currently allopatric conspecific populations clustered with one another, demonstrating the potential of microsatellites for phylogenetic purposes. It appeared that the best resolution was obtained using genetic distance statistics that do not assume any particular microsatellite mutation model. The AFLP phylogeny, based on data from eighteen primer pairs, demonstrated the occurrence of parallel evolution of male colour patterns in different lineages. The origin of a major change in female coloration was less clear but the data suggest that polymorphism for the trait tested may be an ancestral condition.

Acknowledgements

I would like to thank my supervisors Norman Maclean and George Turner for all their help, advice and support during my studentship. Many thanks must go to Mairi Knight and Ole Seehausen who both provided valuable samples and equally valuable advice during my research.

I would also like to express my thanks to a number of people who provided assistance and support; Arati Iyengar, Suzanne Brooks, Aziz Rahman, Sanit Piyapattanakorn, George Harper, Matt Henderson, Mark Dixon, Martin Genner, Martin Taylor and Hala Ayad. Thanks must also go to Tom Kocher and members of his research group, particularly Pat Danley and Craig Albertson, for their hospitality during my visit and for samples and protocols.

The research contained in this thesis was financially supported by a NERC studentship. The Royal Society and NERC provided additional monies.

On a more personal note, I would like to thank my mother for endless support and encouragement, and my father for showing me the true value of 'SGS'.

Nick, what can I say? It's been almost three years – the best things are definitely worth waiting for…..

TABLE OF CONTENTS

TABLE OF CONTENTS ... i
TABLE OF FIGURES ... iv
1 General Introduction and Literature Review ... 1
1.1 Preface ... 1
1.2 Introduction to Molecular Markers ... 1
1.2.1 Allozymes ... 1
1.2.2 DNA Sequence Data ... 3
1.2.3 Other DNA based Markers ... 3
1.2.4 Microsatellites ... 6
1.2.5 AFLPs ... 12
1.3 Speciation Events ... 14
1.3.1 Darwin's Finches ... 14
1.3.2 Baikal Sculpins ... 15
1.3.3 Hawaiian Crickets ... 15
1.4 Speciation in the African Great Lakes ... 16
1.4.1 The mbuna: model taxa for the study of evolutionary processes ... 18
1.4.2 Investigating the evolutionary history of cichlids ... 23
1.4.3 A note on mbuna taxonomy ... 24
1.4.4 Summary ... 28
1.5 Aims and Scope of Thesis ... 28
1.6 References ... 29
2 Improving Microsatellite Based Estimates of Relatedness ... 38
2.1 Introduction ... 38
2.2 Methodology ... 40
2.2.1 Sample Origin and DNA Extraction ... 40
2.2.2 PCR Conditions ... 41
2.2.3 Statistical Analysis ... 42
2.3 Results ... 45
2.3.1 Linkage ... 45
2.3.2 Calculations of Pairwise R ... 45
2.3.3 Rarefaction Analysis ... 46
2.3.4 Simulations ... 46
2.4 Discussion ... 51
2.4.1 Optimisation of extra loci ... 51
2.4.2 Relatedness Estimates ... 51
2.4.3 Conclusions ... 54
2.5 REFERENCES ... 55
2.6 Chapter 2 Appendix ... 58
3 Allopatric Populations and Tests of Mate Choice ... 64
3.1 Chapter Note ... 64
3.2 Introduction ... 64
3.3 Methods ... 67
3.3.1 Experimental Design ... 67
3.3.2 Genetic Analysis ... 68
3.4 Results ... 69
3.4.1 Assortative Mating ... 69

3.4.2 PCR Problems 70
3.5 Discussion 71
3.5.1 Problems with PCR 74
3.6 REFERENCES 76
4 Mate Choice of Cichlid Fish from Lake Victoria 94
4.1 Chapter Note 94
4.2 Introduction 94
4.2.1 Lake Victoria 94
4.2.2 The cichlids of Lake Victoria 95
4.2.3 Aims 97
4.3 Methods 97
4.3.1 Experimental set-up 97
4.3.2 Genetic Analysis 99
4.4 Results 103
4.4.1 Utility of Loci chosen 103
4.4.2 Allele Frequency distributions 103
4.4.3 Mate Choice of P. pundamilia and P. nyererei females 104
4.4.4 Mate Preference of Hybrid females 105
4.4.5 BROOD simulations 105
4.5 Discussion 109
4.5.1 Utility of Microsatellite Markers 109
4.5.2 Non Assortative Mating? 110
4.5.3 Mate Preference of F1Hybrid females 111
4.5.4 BROOD Simulations and Sample Size 112
4.5.5 Conclusions 113
4.6 REFERENCES 114
4.7 Appendix 116
5 Using Microsatellites to Reconstruct the Phylogeny of Lake Malawi Cichlid Fish 132
5.1 Introduction 132
5.1.1 Aims 133
5.2 Methods 135
5.2.1 Populations Included 135
5.2.2 Genetic Distance Measures 136
5.2.3 Tree Construction 137
5.2.4 PCA 137
5.3 Results 139
5.4 Discussion 149
5.4.1 Microsatellites and Phylogenetic Signal 149
5.4.2 'Outlying' Populations 150
5.4.3 Position of Pelagic cichlids 151
5.4.4 Problems with Microsatellite Phylogenies 152
5.4.5 Conclusions 154
5.5 REFERENCES 155
5.6 Chapter 5 Appendix 158
6 Testing parallel evolution of colour patterns in the mbuna using AFLPs 159
6.1 Introduction 159
6.2 Methods 161
6.2.1 Traits and species used 161
6.2.2 Sample collection and extraction 166

6.2.3 AFLP Procedure 166
6.2.4 Visualisation 169
6.2.5 Band Scoring 169
6.2.6 Testing Reproducibility 170
6.2.7 Phylogenetic Analysis 171
6.3 Results 172
6.3.1 Reproducibility 172
6.3.2 Non Random Structure in the data 174
6.3.3 Changes in Peak Height Threshold 177
6.3.4 'Odd' individuals 177
6.3.5 Effect of the number of primer pairs used on tree resolution 178
6.3.6 Tree Topology and Resolution 178
6.3.7 Population Relationships and Evolution of Colour Traits 180
6.4 Discussion 190
6.4.1 Utility of AFLPs 190
6.4.2 Variations in Tree Topology 191
6.4.3 Why have some male colour traits evolved in parallel? 194
6.4.4 Parallel Speciation as well as Parallel Evolution? 195
6.4.5 Conclusions 195
6.5 References 197
6.6 Chapter 6 Appendix 200
7 General Discussion 207
7.1 Future Work 210
7.2 REFERENCES 212

TABLE OF FIGURES

Figure 1.1. Schematic representation of the generation of AFLP markers.. 13
Figure 1.2 Satellite map of East Africa showing the relative locations and sizes of the Great Lakes. 19
Figure 1.3. Phylogenetic tree based on mtDNA data redrawn from Meyer (1993).. 25
Figure 1.4. Phylogenetic relationships of African cichlids based on cytochrome b sequences.. 26
Figure 1.5. The phylogenetic reconstruction of four mbuna genera produced using AFLPs by Albertson *et al* (1999)........ 27
Figure 2.1. The mean and range of R values generated for four relatedness categories 47
Figure 2.2. Deviation from expected R value calculated for known relatives for the four relatedness categories.. 49
Figure 2.3. Rarefaction analysis for 11 loci shows that there is little improvement in relatedness estimation beyond 8 or 9 loci. 49
Figure 2.4. Distributions of 1000 R values for simulated full sibs, half sibs and unrelated pairs. 50
Figure 3.1. a) Male representatives of the Nkhata Bay race (Nkhata Bay population) and the Ruarwe population of *M. zebra*. 66
Figure 4.1. Lake Victoria and surrounding lakes 94
Figure 4.2. Representatives of the species used in this study. a) *P. pundamilia* male, b) *P. nyererei* male. 98
Figure 4.3. Allele frequency distributions for the *P. pundamilia* and *P. nyererei* laboratory populations at a) UNH130 and b) UME003 104
Figure 4.4. Histograms of cumulative values of N* based on 100 resamplings of 100 simulated broods. 106
Figure 4.5. Comparison of behavioural and microsatellite based indices of mate preference in *P. pundamilia*, *P. nyererei* and F1 hybrid females. 112
Figure 5.1. Phylogeny of Malawi cichlids based on a combination of NAD2 and D loop sequences. 134
Figure 5.2 Locations of mbuna populations sampled. Lower map shows relative locations of populations within Nkhata Bay 138
Figure 5.3. Majority rule NJ tree constructed using the delta mu distance measure of (Goldstein *et al* (1995b). 141
Figure 5.4. Majority rule NJ tree constructed using Nei's (1972) distance measure. . 142
Figure 5.5. Majority rule NJ tree constructed using the Dps distance measure of Stephens *et al* (1992 143
Figure 5.6. Majority rule NJ tree constructed using the Dc distance measure of Cavalli-Sforza and Edwards (1967). 144
Figure 5.7. Allele frequency distributions for the two populations of *M. zebra* species 'gold' 145
Figure 5.8. Unrooted representation of the majority rule NJ tree produced using the Dc distance measure. 146
Figure 6.1. Sampling locations for populations used in this study. 162
Figure 6.2. Photographs of representatives of the mbuna populations included in this study grouped by the traits being tested 163
Figure 6.3 Sequence of the adapters ligated to the cut DNA fragments 168
Figure 6.4. Number of bands detected per primer pair in each of the three data sets. 173

Figure 6.5. Distribution of the lengths of 100000 randomly generated trees for various numbers of populations 175
Figure 6.6. Effect of addition of successive loci on mean bootstrap value 179
Figure 6.7. NJ tree constructed using the distance of Link *et al* (1995). 182
Figure 6.8. NJ tree constructed using the distance of Nei and Li (1979).. 183
Figure 6.9. NJ tree constructed using the distance of Link *et al* (1995 184
Figure 6.10. NJ tree constructed using the distance of Nei and Li (1979). 185
Figure 6.11. NJ tree constructed using the distance of Link *et al* (1995) 186
Figure 6.12. NJ tree constructed using the distance of Nei and Li (1979).. 187
Figure 6.13. NJ tree of all mbuna populations based on peaks scored using Genotyper (manual alignment) and the genetic distance measure of Nei and Li (1979).. 188
Figure 6.14. NJ tree of all mbuna populations based on peaks scored using Genographer and the genetic distance measure of Nei and Li (1979). 189
Figure 6.15. NJ tree constructed using the distance of Nei and Li (1979).. 201
Figure 6.16. NJ tree constructed using the distance of Nei and Li (1979).. 202
Figure 6.17. Nei 400. NJ tree constructed using the distance of Nei and Li (1979). .. 203
Figure 6.18. NJ tree constructed using the distance of Nei and Li (1979). 204
Figure 6.19. NJ tree based on data scored using Genographer and the distance measure of Link *et al* (1995).. 205
Figure 6.20. NJ tree based on data scored using Genographer and the distance measure of Link *et al* (1995).. 206

1 General Introduction and Literature Review

1.1 Preface

The genetic component of biological diversity has attracted great interest in recent years, and a wide range of molecular methods has been developed to investigate variation at a genetic level. The development of these methods has permitted the investigation of biological and ecological phenomena that would have been impossible using conventional techniques. Current applications of these techniques include analysis of breeding behaviours, phylogeography and population dynamics in a wide range of organisms. When combined with field and laboratory observational and behavioural studies, they have provided a powerful tool for revealing the evolutionary biology of many taxonomic groups, including the prolifically speciating cichlid fishes of the African Great Lakes. This group of organisms represents a fascinating example of explosive speciation and adaptive radiation, and this thesis describes how molecular markers can be applied to problems of relatedness, mate choice and phylogenetics in order to infer evolutionary mechanisms. This introduction will describe the major classes of molecular markers and indicate the types of investigations to which they are best applied, and provide a brief description of other examples of notable species radiations. The introduction concludes with a summary of the current state of knowledge regarding the evolutionary ecology of the cichlids fishes of the African Great Lakes.

1.2 Introduction to Molecular Markers

1.2.1 Allozymes

Allozymes were one of the earliest markers available for the study of individuals and populations. They are allelic variants found at a functional protein (usually metabolic enzyme) locus, and are visualised through differential mobilities of alleles during starch, acrylamide or cellulose acetate gel electrophoresis. They are co-dominant markers – heterozygotes can be readily distinguished as such by electrophoresis– and are normally inherited in a consistent Mendelian manner. Originally they were used to

quantify genetic variation present in humans (Harris, 1966) and in *Drosophila* (Johnson *et al*, 1966). The results of these studies were surprising; 30% of surveyed loci were polymorphic. The amount of variation present allowed allozymes to be used to study large number of genetic and ecological problems. Studies of variation within and between populations heralded the development of statistical estimators of population genetic parameters, for example the F_{ST} group of statistics as defined originally by Wright (1951). These statistics are still in use today, and have also been adapted to data from more recently developed markers. The co-dominance and predictable inheritance of allozymes allows them to be applied to studies of mate choice, for example paternity exclusion or the detection of multiple paternity. A recent example of this kind of application is the study of Webb *et al* (1995) who used enzyme polymorphisms to assess the genetic structure of wild rabbit (*Oryctolagus cuniculus*) populations. Evolutionary relationships between taxa can also be estimated using allozymes, as carried out by Sage *et al* (1984) who deduced broad scale relationships between 13 species of African cichlid fish using 10 loci. There are a few drawbacks associated with allozymes as molecular markers. The first is the need to preserve and store samples so that the proteins do not degrade – allelic differences are only detectable if the protein remains in an active state. This requires immediate storage of samples in dry ice or at temperatures as low as –80°C, a problem when large numbers of sample are collected during fieldwork in remote locations. Another problem may be finding enough variable loci in the organism of interest as 10-20 appropriately variable loci are usually required for statistical analysis (Parker *et al*, 1998). Variation at allozyme loci is often low when compared to that detected by more recently developed DNA markers. This is because allozymes represent functional enzymes, and any mutations in amino acid sequence which lead to reduced functionality must surely be heavily selected against (Parker *et al*, 1998). In addition, the degenerate nature of the genetic code means that some point mutations in DNA sequence make no difference to the amino acid sequence of the protein (so called 'silent' substitutions). This means that quite often, not enough variation is detectable to distinguish between individuals, populations or even species. Some allozyme loci have been revealed to be under direct selective pressure, for example Watt *et al* (1993) found that temperature tolerance was directly linked to variation at the PGI locus in *Colias* butterflies. Instances such as this allow for interesting studies of ecological genetics, but for the purpose of studying divergence among populations many statistics assume the loci involved to be selectively neutral.

1.2.2 DNA Sequence Data

Unlike allozymes, DNA (or RNA) sequence data is capable of providing information at all levels of variation, both coding and non-coding. However, the cost and effort of sequencing dictates that only small, specific areas of the genome are sequenced, and only relatively few individuals from each taxon are used (Avise, 1994 p90, Hillis *et al*, 1996 p339). In some investigations, information is required from many loci across the genome, and sequence data from a single gene would not be appropriate. However, sequencing is an extremely popular method for collecting data for phylogenetic and phylogeographic analysis (Hillis *et al* 1996, p321). Mutation rates are not homogenous across the genome, and this allows the researcher to select an appropriate region to sequence, depending on the evolutionary timescale under consideration. For example, higher animal mitochondrial DNA mutates on average 5-10 time faster than single copy nuclear DNA (Avise, 1994 p101) making it suitable to study population level differences and recent divergence between species. Other more slowly mutating regions have also been utilized for studies of deep divergence, for example Hamby and Zimmer (1992) examined relationships between the major plant clades using nuclear 18S and 26S ribosomal RNA genes.

1.2.3 Other DNA based Markers

A number of other molecular markers have been developed which do not require the cost and effort associated with sequencing. These also have the potential advantage of sampling variation across many areas of the genome instead of a single gene or two. Such markers have been developed both prior to the advent of wide scale sequencing (for example DNA hybridisation and RFLPs) and subsequent to it (multilocus fingerprinting, RAPDs, microsatellites). A major development in the history of molecular markers was the advent of PCR, the polymerase chain reaction (Saika *et al*, 1985, Mullins and Faloona, 1987). This enabled defined regions of the genome to be amplified without the need for lengthy cloning and hybridisation protocols. PCR provides enough genetic material for subsequent analysis, for example with RFLPs (see below), and is also fundamental to many of the more recently developed markers. A brief description of each method is given below along with the types of investigation it

is best suited to. Microsatellites are described in greater detail as they have been used in four out of the five experimental chapters in this thesis.

1.2.3.1 DNA hybridisation

Originally used by studies such as that of Britten *et al* (1974), DNA hybridisation provides a measure of how similar the single or low copy number fractions of the genomes of two species are. Double stranded DNA from two different organisms is mixed and then denatured. The rapidly reannealing repetitive fraction is removed and discarded. Remaining single strands (representing low copy number or single copy number DNA) are radioactively labelled and then mixed with the DNA from another species. Heating the resultant duplex will melt it; the temperature at which it melts is dependent on the degree of similarity between the DNA of the two species and the base composition of the strands. The resulting data is in the form of 'thermal elution profiles', which give indications of the genetic distance between two species. A notable example of a study using DNA hybridisation is that of Sibley and Ahlquist (1990), which investigated relationships between 1700 bird species. A drawback of this kind of marker is that the kinetics of reassociation are not yet fully understood because multiple factors such as genome size, copy number, fragment size and base composition exert an influence (Werman *et al*, 1996). The fact that data from hybridisation experiments is in the form of distance data and not character states may limit subsequent phylogenetic analysis.

1.2.3.2 Restriction Fragment Length Polymorphisms (RFLPs)

This class of molecular marker utilises enzymes (restriction endonucleases), which occur in bacterial cells. These enzymes act to protect the host from the invasion of foreign DNA. Type II restriction endonucleases cut double stranded DNA at specific 4, 5, or 6 base recognition sequences. Restriction enzymes may be used singly or in combination to produce sets of cut DNA fragments (the digestion profile). These fragments can be visualised by gel electrophoresis followed by chemical staining or radioactive hybridisation (Southern blotting). Variation in the digestion profiles between individuals is caused by either the gain or loss of a restriction site or by

changes in the length of the DNA between the two sites. Different classes of DNA may be analysed by this technique, for example the entire mitochondrial or chloroplast genome can be subjected to restriction analysis (published studies include those of Avise *et al*, 1987 and Seutin *et al*, 1993). It is also possible to use restriction enzymes to analyse single copy nuclear DNA and particular gene families such as ribosomal RNA genes.

1.2.3.3 Multilocus Fingerprinting

Minisatellites are repeats of generally 16-64 base pairs that are scattered throughout the genomes of most organisms (Avise, 1994 p79). They are highly variable and most mutations are thought to be caused by strand slippage or unequal crossover during replication (Jeffereys *et al*, 1988). Minisatellites are usually analysed by cutting genomic DNA with restriction enzymes, electrophoresis of fragments and then Southern hybridisation with a suitable probe (e.g. that of Jeffereys, 1985). The same probe may be effective in many different taxa (Avise, 1994 p80). This process results in a profile of 10-30 bands per individual (Parker *et al*, 1998). Due to the polymorphic nature of minisatellites, such profiles are powerful tools for individual identification and parentage assignment. However their use in population genetics is somewhat limited because there is no way of determining which bands represent alleles at a particular locus, which is a requirement of many statistics. Another consideration is that profiles that result from runs on different gels may not be directly comparable. An alternative approach is to develop locus specific probes for minisatellites that would enable allelic states to be determined. However, as this method is labour intensive (involving library construction, sequencing of probes and sequential hybridisations), it is now considered much more feasible to use microsatellites (Parker *et al*, 1998) which are described in detail below.

1.2.3.4 Randomly Amplified Polymorphic DNA (RAPDs)

Used originally by Williams *et al* (1990), RAPDs are another class of marker which samples variation across the genome. In theory, they are relatively straightforward to use as they are based on the polymerase chain reaction. A single short primer (around 10 bases) of random sequence is used; the advantage of this is that no lengthy primer

design stage is required for new taxa. The resulting amplification profile is then visualised by agarose gel electrophoresis. Variation between individuals will cause the gain or loss of a fragment, although heterozygotes may not be distinguished from homozygotes as most RAPD markers are not co-dominant (Parker, 1998). However, they have proved useful in aspects of molecular ecology such as paternity assignment (Hadrys *et al*, 1993) and population structure (Jacobsen *et al*, 1996). The major problem associated with RAPDs is a lack of reproducibility – PCR artefacts and contamination are common problems (Allender, MRes dissertation, University of Warwick, 1998). In addition, some studies have reported problems with reproducibility of profiles between (Jones *et al*, 1997) and even within (McEwan *et al*, 1998) laboratories. They should therefore be approached with caution when selecting appropriate markers for study.

1.2.4 Microsatellites

Microsatellites are regarded as one of the most powerful genetic markers currently available (Goldstein and Pollock, 1997; Jarne and Lagoda, 1996). They are comprised of short segments of tandemly repeated DNA, the repeated unit being anywhere from one to six bases long. A microsatellite locus is usually under one hundred repeat units in length (Jarne and Lagoda, 1996). Alleles at a particular locus will vary in the number of repeat units present; high levels of variability can often be found with as many as 40 alleles occurring at a particular locus (Queller *et al*, 1993). Variations in the number of repeat units mean that the microsatellite differs in length. Allelic differences can therefore be resolved using gel electrophoresis. Microsatellites also exhibit co-dominance, which means that heterozygotes can be distinguished from homozygous individuals when scoring alleles.

Microsatellites are found throughout the eukaryotic genome, although their distribution and frequency varies amongst different taxa (Lagercrantz *et al*, 1993). Although they are located in both coding and non-coding areas of the genome, it is widely believed that microsatellite DNA is not expressed. This is due to its location outside of exons or because repeated regions appear to be spliced out during the transcription process. If microsatellite DNA is not expressed as protein, then it can be assumed to be selectively neutral, with no difference in selective advantage between alleles at a locus. However,

certain evidence has been accumulated to challenge this assumption, this will be discussed further below.

The origin of such short repetitive sequences of DNA remains unclear. It has been suggested that microsatellites arise and mutate due to the 'slippage' of the polymerase enzyme during replication. This mechanism, which was demonstrated *in vitro* by Schlotterer and Tautz (1992), would cause the repeated sequence to grow by one or two repeat units per slippage event. However, other mechanisms such as gene conversion and unequal crossover may be responsible for rarer, larger scale changes in repeat number (Estoup and Angers, 1998). Whatever the mutation mechanism, microsatellites are known to have a relatively high mutation rate. In humans, it has been found that 1×10^{-3} to 1×10^{-4} mutation events occur per locus, per generation (Weber and Wong, 1993). This compares with estimates of point mutations in non-microsatellite sequences of around 1×10^{-9} to 1×10^{-10} events per base per generation. Such a high mutation rate makes them ideal markers for studying recently diverged populations and species. Variation in DNA accrued between closely related groups may often not be revealed by molecular markers with a slower rate of mutation.

1.2.4.1 Statistics for microsatellites

In order to use data obtained from microsatellite DNA for the investigation of population attributes such as genetic structure, it is vital that an appropriate statistical method is used in the analysis. There are various statistical methods available in the literature that provide estimates of genetic distance between two populations. Some have been developed specifically for microsatellite data, whilst others were developed to analyse data from other molecular methods. It should be borne in mind that microsatellites have mutation patterns that are not shared by other molecular markers. These characteristics are especially pertinent when attempting to model the mutation and evolution of microsatellites. The Infinite Alleles Model (IAM, Kimura and Crow, 1964) assumes that at a particular locus, any allele is capable of generating any other allelic state in a single mutation. However, it has been noted and demonstrated that microsatellite mutation frequently occurs as a result of slippage of the polymerase enzyme during DNA replication (Schlotterer and Tautz, 1992). This results in either the gain or the loss of a single repeat unit. Ohta and Kimura's (1973) Stepwise Mutation

Model (SMM) has been assumed to apply to microsatellites (Shriver *et al*, 1993; Valdes *et al*, 1993). In addition to the mechanism of mutation, the SMM also assumes that there is no upper limit to the number of repeats possible at a particular locus.

In a recent review of genetic distance measures, Goldstein and Pollock (1997) noted that due to the different assumptions of these models, statistics based on one or the other perform very differently when applied to microsatellite data. A distance measure such as Nei's standard distance (based on the IAM) will perform in a non-linear and inaccurate manner when used to analyse data sets obtained from microsatellite DNA where the loci concerned mutate in a stepwise manner. Distance measures based on the SMM may therefore be more appropriate, and several have been developed for example $(d\mu)^2$ (Goldstein *et al*, 1995b). A fixation index equivalent to F_{ST} has also been developed which incorporates the SMM and is applicable to microsatellite data (R_{ST}, Slatkin, 1995). Other distance measures include ASD (Goldstein *et al*, 1995), and Shriver *et al*'s D_{SW}. All three of these distance measures make use of size difference between alleles. Another group of distance statistics do not assume any particular evolutionary model to be true. These include D_C (Cavalli-Sforza and Edwards, 1967), D_A (Nei *et al*, 1983) and D_{AS}, the allele sharing distance developed by Stephens *et al* (1992). Other methods of obtaining distance measures have also been developed. Di Rienzo *et al* (1994) found that a two stage mutational model based primarily on single step mutations, but which allows for larger scale changes in repeat number was successful at predicting allelic variation in human populations. Paektau *et al* (1997) tested several distance measures on data sets from several Arctic bear populations. It was found that distance measures developed specifically for microsatellites were less informative than conventional distance measures at finer scales. The authors suggested that this was due to the higher variance of microsatellite distance measures. However, at larger evolutionary scales, all distance measures performed poorly when used to analyse microsatellite data. The authors attributed this to uneven and constrained mutation of microsatellites. It is thus apparent that not all statistical treatments are applicable to all sets of loci, and the appropriate genetic distance statistic must be chosen based on the characteristics of the loci concerned and the evolutionary timescale under investigation.

1.2.4.2 Problems associated with microsatellites

Although large numbers of studies have been published which use microsatellites, it has become apparent that current understanding of mutational processes and the development of suitable statistics limit interpretation of microsatellite data. For example Feldman *et al* (1997) discuss the statistical consequences of the fact that allele sizes have an upper and lower boundary. Several complications have also been reported in the literature that can affect the validity of analyses.

Several authors have reported the phenomenon of size homoplasy. This occurs where alleles share the same size as determined by electrophoresis, but not the same sequence. It is possible to detect several alleles that are identical in size, but not identical by descent. Orti *et al* (1997) report such an occurrence in a CA repeat isolated from horseshoe crabs (*Limulus polyphemus*). Variations in the flanking regions (a non repeat region) of the microsatellite were found to contribute to size alterations. The size of the allele was therefore a poor indicator of phylogenetic relationships between them. Angers and Bernatchez (1997) reported similar findings for a microsatellite locus originally isolated from brook char (*Salvenius fontinalis*).

In order to accurately reflect between and within population differences, the molecular markers used must be selectively neutral i.e. a particular allele at a particular locus does not confer any selective advantages. As discussed previously, this was assumed to be true for microsatellites since it was thought that they are not involved in gene expression. However Streelman *et al* (1998) report that the sequence of a microsatellite flanking region shows similarity to RAS-specific guanine nucleotide releasing factor. This hints at a function for this particular locus. It is also known that large repeat numbers (e.g. trinucleotide repeats) at certain human loci are associated with various genetic disorders.

Mutations in primer binding sites can cause the failure of a particular allele to amplify. Such occurrences are known as null alleles, and they can cause considerable difficulty when estimating the heterozygosity of a population, and also result in false exclusions in paternity testing. Individuals that are heterozygous for one amplifying allele and one null allele will be scored as homozygous. The molecular changes behind a null allele

found in the white sands pupfish (*Cyprinadon fularosa*) was described by Jones *et al* (1998). In fact several alleles had been failing to amplify. Null alleles can be identified and assessed by sequencing and then redesigning the primers in order to obtain a successful PCR.

The most frequently observed PCR artefact of microsatellites is stutter, where a number of PCR products are produced, each one repeat unit shorter than the preceding one. It is thought to arise by slippage of the polymerase enzyme during PCR, or by incorrect annealing of primers. It can be a problem if stutter products are present in the same quantity as the true alleles, or if stutter products are larger than the true allele. However, in most cases stutter forms an interpretable pattern that does not cause problems with allele identification.

Other PCR artefacts that can cause problems are allelic dropout and the plus A activity of Taq polymerase. Allelic dropout is a consequence of the fact that shorter templates will always amplify much more effectively than longer ones. The longer allele of a heterozygote will always be visualised less intensely than the shorter. This was effectively demonstrated by Wattier *et al* (1998) who created artificial heterozygotes by mixing two haploid DNA samples. They suggest that the correct PCR parameters need to be carefully determined to minimise this effect. Taq polymerase will sometimes add a non-templated A residue to the end of a PCR product. This can cause difficulties in sizing alleles, particularly if stutter is present as well. A series of PCR products are generated, each differing in size by only one base. This phenomenon has been extensively studied, and several suggestions of how to promote or prevent it have been published (Brownstein *et al*, 1996, Magnuson *et al*, 1996). In practice, it is easier to promote plus A, and the most successful method appears to be the addition of a seven base tail to the reverse primer. Altering the thermal profile of the PCR can also help.

1.2.4.3 Applications of microsatellites

Microsatellites have been applied to a wide variety of ecological phenomena, and have also been used to examine evolutionary relationships. They appear to have been particularly useful for detecting genetic structure within populations. One application of this is the management of fish populations. For example, the genetic structure of

populations of species of Pacific salmon was assessed using microsatellites (Olsen *et al*, 1996; Wenburg *et al*, 1998 and Wenburg *et al*, 1996). Populations were found to be structured, even within a single river system, with obvious implications for environmental alterations to spawning areas and species conservation.

The usefulness of relatedness estimates generated from microsatellite data was tested by Blouin *et al* (1996) who were able to distinguish between unrelated and full sib pairs of individuals with at least 97% accuracy. Half sibs and unrelated individuals could be distinguished in more than 80% of samples. The results indicated the usefulness of data obtained from microsatellites for examining relatedness in wild populations. Taylor *et al* (1997) used eight to nine microsatellite loci to determine relatedness amongst individual hairy nosed wombats (*Lasorhinus krefftii*) occupying the same burrow. The molecular data was gathered in order to test the hypothesis that dispersal in this species is female biased and occurs after breeding. Microsatellite data were similarly used by Surridge *et al* (1999) to examine the fine scale genetic structure present within a colony of wild rabbits (*Oryctolagus cuniculus*). The results indicated the operation of male biased dispersal since the relatedness of males within social groups and within the colony as a whole was lower than that of females.

Microsatellites have also proved to be useful for certain types of phylogenetic reconstruction. Goldstein and Pollock (1997) note that there are several problems associated with the use of microsatellite data for phylogenetic purposes. Due to extraordinarily high mutation rates, over long periods of evolutionary time, microsatellite repeats can degrade, and often are no longer recognisable as being homologous with a putative ancestral state. Degradation of microsatellites may occur through interruption of the repeat region, which then acts to prevent further polymerase slippage. The repeat area can then be lost through substitutions and insertions/deletions over the course of evolutionary time (Taylor *et al*, 1999). There also appears to be an upper limit to the number of repeats possible at a microsatellite locus, and it seems that alleles with high numbers of repeats mutate faster than those with lower numbers (Feldman *et al*, 1997). Thus modelling the mutation and evolution of microsatellites is not straightforward. However, Rico *et al* (1996) report that flanking regions of microsatellites were conserved between fish species, which diverged approximately 470 million years ago. However, whether informative variation is still present after

such a long period of time remains to be seen. Feldman *et al* (1997) suggested that for phylogenetic analyses, microsatellite loci should be chosen based on a relatively small mutation rate (to prevent complete degradation of the locus) and a large range in allele size. Microsatellites appear to be useful for reconstructing phylogenies of closely related species, or of groups which speciated in a very short period of time. The phylogeny of Darwin's finches of the Galapagos Islands has been reconstructed using allozyme and mitochondrial data. Petren *et al* (1999) utilised sixteen microsatellite loci to produce a phylogeny that was broadly consistent with the previous two. However, the microsatellite data revealed a deep division between two finch populations that had previously been placed in the same genus. The authors noted that this split was made apparent by the rapidly mutating nature of microsatellites, and that they were ideal markers for reconstructing the phylogeny of recently evolved, closely related taxa.

The above references provide a few examples of the kinds of evolutionary and ecological questions which microsatellite DNA has been used to answer. They are versatile markers, which can be used at different genetic scales, from answering questions about the pedigree of an individual to looking at the evolutionary relationship between different taxa. The attributes of microsatellite DNA described above (co-dominance, rapid mutation, high levels of polymorphism) make it an ideal tool to study relatively young, closely related groups of species. However, one should be aware of the potential problems (as discussed earlier) associated with this class of molecular marker.

1.2.5 AFLPs

First described by Vos *et al* (1995), AFLPs involve the use of restriction enzymes and adapter molecules to provide annealing sites for PCR primers (see Figure 1.1). The name refers to the amplification of restriction fragment length polymorphisms, although fragments are scored on a presence/absence basis and not differences in size. In brief, genomic DNA is cut with two restriction enzymes, a rare cutter and one that cuts more frequently (for example *Mse*I and *Eco*RI). Adapters specific to each enzyme recognition site are ligated to each fragment. The sequence of these adapters, combined with the sequence of the restriction enzyme recognition site serve as a primer-binding site for PCR. Altering the last nucleotides at the 3^1 end of each primer may generate

different sets of fragments. Mutations in either the restriction sites or in the region to which the selective sequence of each primer binds will prevent amplification of a particular fragment. The final data are therefore in the form of band presence and absence information for each individual tested. These markers have the advantage of all PCR based markers in that they only require small amounts of template DNA but, in addition, no sequence information is required prior to typing. They therefore do not have the lengthy development phase associated with microsatellites (Mueller and Wolfenbarger, 1999).

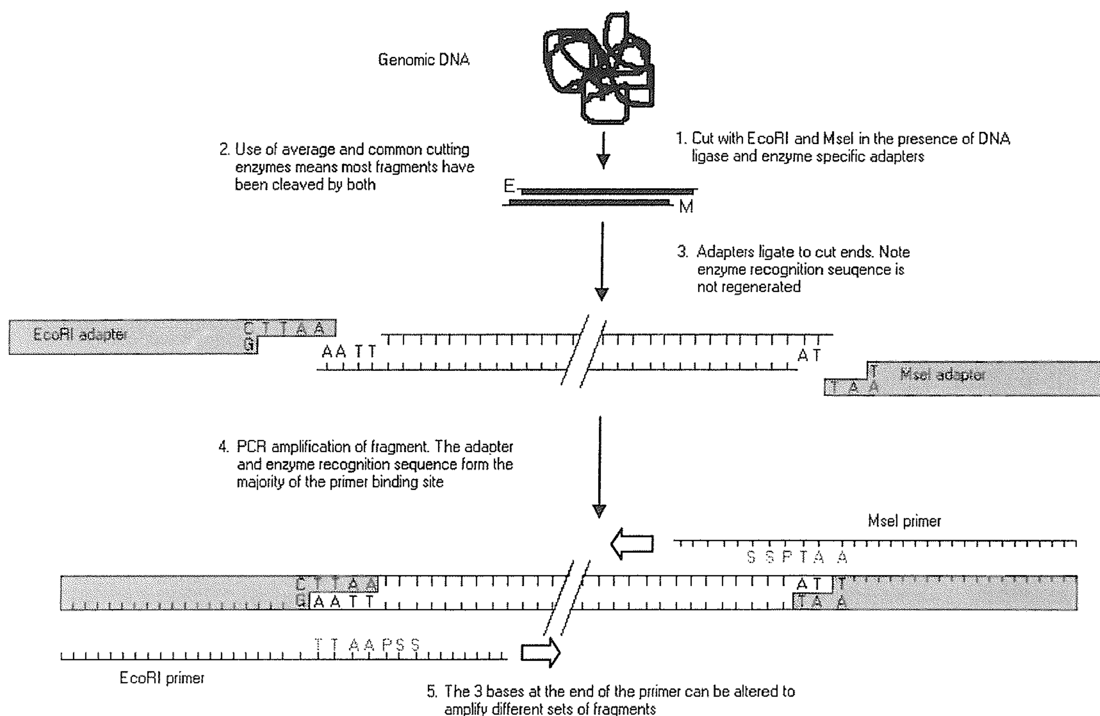


Figure 1.1. Schematic representation of the generation of AFLP markers. The enzyme specific adapters have a point base mutation compared to the enzyme recognition sequence. This ensures fragment/adapter ligations are not re-cut while fragments which ligate to one another are recleaved by the enzyme. Note that different combinations primers will amplify different subsets of fragments depending on the nucleotides present at the pre-selective (P) and selective (S) positions.

AFLPs survey variation across the whole genome, and thus have been particularly well used in commercial species where genome and quantative trait loci mapping are desirable (for example *Brassica oleracea* by Sebastian *et al*, 2000, *Oreochromis niloticus* by Kocher *et al*, 1998). Other applications include species and strain

identification in microbiology (Savelkoul *et al*, 1999). The use of AFLPs in the study of population differentiation has also proved successful in both cultivated (e.g. sunflowers, Quagliaro *et al* 2001) and non- cultivated (e.g. *Arabidopsis*, Sharbel *et al*, 2000) species. However, there appear to be relatively few published studies regarding the use of AFLPs to investigate the population genetics or phylogenetics of animal species. AFLPs appear to be particularly applicable to the study of closely related species where surveys of genome wide variation would provide the best phylogenetic resolution. These have included investigation of genetic diversity and species boundaries within Hawaiian crickets (Parsons and Shaw, 2001) and intraspecific relationships within a snake species (Giannasi *et al*, 2001).

1.3 Speciation Events

There are several well-studied examples of contemporary taxa that have undergone major speciation events. Terrestrial examples include Darwin's finches (14 species, Petren *et al*, 1999) and Hawaiian *Drosophila* (800 species, Powell and Desalle, 1995). Examples of speciation events in lakes are Lake Baikal cottoid fishes (31 species, Sidelera, 2000), thiarid gastropods from Lake Tanganyika (70 species, West and Michel, 2000) and the cichlid fish of the African Great Lakes. The age, magnitude and biogeographical context of three well described speciation events are described in further detail below.

1.3.1 Darwin's Finches

The radiation of finches ('Darwin's Finches) on the Galapagos Islands is a particularly well-known and well-studied example of adaptive radiation. The Galapagos Islands are an archipelago of 19 islands located 960km west of the coast of Ecuador. Currently, 14 species of finch are recognised (Petren *et al*, 1999), and the radiation is thought to have occurred in less than three million years (Grant, 1993). The most notable feature of the radiation is that the variation in beak and body shape and size exhibited by the Galapagos finches is more characteristic of divergence between families of birds (Petren *et al*, 1999). Molecular data do not indicate that any extreme bottleneck has taken place, suggesting that the effective breeding size of the founding population was

not smaller than 30 individuals (Vincek *et al*, 1997). Although this study could not rule out the possibility of multiple colonisations, the authors felt that this was unlikely given the distance between the islands and the mainland. The finch radiation appears to be monophyletic as supported by the microsatellite phylogeny of Petren *et al* (1999). Hybridisation between certain finch species is known to occur in the wild (Grant, 1993). However, the level of gene flow between species is not enough to blur phylogenetic relationships within the group; hybridising populations do not form phylogenetic clusters as they remain genetically closer to conspecifics rather than the sympatric heterospecifics with which they occasionally interbreed (Petren *et al*, 1999).

1.3.2 Baikal Sculpins

An example of lacustrine speciation may be seen in Lake Baikal. Located in a rift valley in Siberia, it is one of the oldest and deepest lakes in the world. The age of the lake is estimated to be 20Myrs (Martens, 1997). It has a maximal depth of around 1600m, and 80% of its surface area has a depth of greater than 250m. Unusually, the deeper layers of the lake are relatively well oxygenated which has allowed a diverse abyssal fauna to develop (Weiss *et al*, 1991). Of the 1825 species present, 54% are endemic (Martens, 1997). There are 52 species and subspecies of fish native to Lake Baikal (Sideleva, 2000). Thirty-one of these taxa belong to a single sub-order, the Cottoidei, and these fish occur at all depth habitats within the lake. The age of the radiation has been estimated to be 4.9Myrs based on sequence data from the rodopsin gene (Hunt *et al*, 1997). The same study also indicated the monophyletic nature of the Baikal cottoids – speciation must have occurred solely within the lake. Hunt *et al* (1997) suggest that adaptation to different depth habitats within the lake could have been one of the causal mechanisms of this radiation.

1.3.3 Hawaiian Crickets

The crickets (family Gryllidae) of the Hawaiian archipelago provide an example of an invertebrate species-rich radiation. Members of this radiation are flightless, with retention of the forewings as sound producing organs (Shaw, 1996). The radiation is thought to be approximately 2.5Myrs old and comprises 206-210 species (Otte, 1989). Most species are endemic to only a single island, although genera have an archipelago-

wide distribution (Shaw, 1996). Sympatric species may be morphologically very similar and only distinguishable by the male song type (Otte, 1989). Species boundaries based on song phenotype are supported by an AFLP based phylogenetic study (Parsons and Shaw, 2001). Speciation in this group is thought to have occurred in an intra-island manner; the geology of the islands (volcanoes and high rocky ridges) acts to reduce gene flow between populations, and linear valleys and lava tubes increase the magnitude of isolation by distance.

1.4 Speciation in the African Great Lakes

It appears that islands and lakes provide the necessary biogeographical environment in which large-scale speciation events may occur (Martens, 1997). Endemic fauna radiations in ancient lakes have provided vital clues about the evolutionary mechanisms causing and maintaining species divergence. The African Great Lakes (Figure 1.2) contain several species rich groups both vertebrate and invertebrate. The cichlid fish of lakes Malawi, Tanganyika and Victoria, provide an astounding example of explosive speciation. From data provided in Table 1.1, one can immediately see that hundreds of species of cichlids are present in each lake. The species flocks of Lake Malawi and Lake Victoria were thought to have arisen from a single ancestral lineage (Meyer *et al*, 1990). However recent studies have suggested that while part of the Lake Victoria flock is monophyletic and evolved within the Lake basin, other clades within the flock probably arose elsewhere in the region (Booton *et al*, 1999; Nagl *et al*, 2000). Despite this, the question that must be asked, therefore, is how has such a large number of species arisen in such a short period of evolutionary time? Owen *et al* (1990) note that endemic colour forms are present in the south east arm of Lake Malawi; an area that they suggest was dry as little as 250 years ago. The fact that the colour morphs concerned are not found anywhere else in the lake suggests that they must have diverged from other morphs after the lake level rose.

There has been much speculation over the predominant evolutionary mechanism of the cichlid radiation. A suggestion that the species richness of Lake Malawi was the result of multiple colonisations was refuted by studies of mitochondrial DNA (Meyer *et al*, 1990; Meyer, 1993, see Figure 1.3), which found the Malawi radiation to be

monophyletic and to have diverged from the ancestral lineage as recently as 700,000 years ago. Suggestions that species evolved in lagoons around the periphery of the lake (the 'Lake Nabugabo scenario – see Greenwood, 1965, cited in Turner, 1999) have also provided much debate. However, this scenario may not be applicable to all the lakes as Owen *et al* (1990) point out that the physical characteristics and water level fluctuations of Lake Malawi mean that peripheral lagoons do not exist for a long enough period of time to be of evolutionary significance. Water levels in all the Great Lakes have fluctuated, sometimes drastically, in the course of their history. Investigations of these changes rely on the examination and radiocarbon dating of sediment cores from various parts of a lake, but sedimentation rates may not be constant and radiocarbon dating is always associated with some degree of error (Kornfield and Smith, 2000). The magnitude and timing of all lake level fluctuations may therefore never be determined with certainty. In general, there appears to have been a major low stand (600m lower than present) of Lake Tanganyika approximately 200,000 years ago, and three more recent low stands of 160-300m (reviewed in Kornfield and Smith, 2000). The geology of the Tanganyikan basin means that major drops in water level effectively split the lake into a number of smaller basins, and these events may explain the current distribution of mitochondrial DNA haplotypes of some of the cichlids present (Sturmbauer *et al*, 2001). Lake Malawi also underwent a 400m decrease in level 18,000 to 10,700 year ago (reviewed in Sturmbauer *et al*, 2001). Owen *et al* suggest that the southern end of Lake Malawi was dry approximately 200 years ago. This means that several endemic cichlid colour morphs must have diverged since the lake rose. There appears to have been a major drop in the level of Lake Victoria 17,000 to 15,000 years ago (Johnson *et al*, 1996). Some authors maintain this event entailed complete desiccation of the lake and surrounding region and go on to suggest that the 200 or so species of cichlids currently or recently occupying the lake must have evolved since then (for example Nagl *et al*, 2000). This hypothesis is controversial as some authors have insisted some pools must have remained in the region and acted as refuges, as Lake Victoria contains a number of non-cichlid endemics, and it would be surprising to some if these groups had also undergone such rapid speciation and evolution (Fryer, 2001; Kornfield and Smith, 2000). However, lake level fluctuations undoubtedly provide the potential for divergence by isolating and then mixing populations. A study of the mtDNA control region in various species of cichlids from each of the three Great Lakes found identical haplotypes in populations

currently separated by large geographic distances, which indicated that these populations had been in contact relatively recently (Sturmbauer *et al*, 2001).

Some authors have pointed out that cichlids possess a 'key innovation' in the form of a set of pharyngeal jaws which would allow flexibility in food intake (e.g. Greenwood, 1984). It is argued that these jaws have allowed diversification in diet, and have thus been responsible for the three adaptive radiations of cichlids seen today in the Great Lakes. However, as Turner (1999) points out, cichlids are not the only group of fish possessing pharyngeal jaws, and also not all groups of cichlids have speciated so profusely. There is a high level of diversity within the East African cichlids in feeding mode, ranging from grazing and plankton feeding to scale biting and paedopahgy (Meyer *et al*, 1990). In a recent phylogenetic study, Albertson *et al* (1999) suggest that divergence in feeding mechanism played an important early role in divergence within one of the clades of Lake Malawi cichlids, but other factors must be responsible for recent and ongoing speciation within the group.

1.4.1 The mbuna: model taxa for the study of evolutionary processes

The magnitude and diversity of the species radiations in all three lakes mean that in order to understand the evolutionary forces at work it is necessary to focus on model taxa. One particularly well-studied group of Malawi cichlids is the mbuna. The mbuna are a group of fish occupying the rocky shore habitat. They generally feed on the encrustations of algae and small animals contained within them (collectively termed 'Aufwuchs') present on the surface of the substrate, although some feed on plankton in the water column (Ribbink *et al*, 1983). The mbuna clade exhibits a high degree of sexual dichromatism, males having bright breeding colours and females usually being more cryptic. Male mbuna are highly territorial and females visit male territories in order to spawn. Like all Malawi cichlids, the mbuna are maternal mouthbrooders (there is no larval dispersal phase in the life history of these fish). All of these factors, combined with a lack of ability to equilibrate to changes in depth, contribute to the generally poor dispersal ability of this group (Ribbink *et al*, 1983). This lack of dispersal ability has led various authors to suggest that mbuna populations occupying a particular rocky habitat patch may be effectively isolated from neighbouring patches, due to the apparent inability of the fish to cross deep channels or areas of sandy

substrate (Fryer and Iles, 1972; Ribbink *et al*, 1983). Mbuna male breeding colours vary between geographic regions (Ribbink *et al*, 1983). Often, there is no discernible morphological variation between geographic races (or 'colour morphs'), only differences in the colour patterns of male breeding dress. The wealth of mbuna variation in colour pattern led

Figure 1.2 Satellite map of East Africa showing the relative locations and sizes of the Great Lakes. A – Lake Victoria, B- Lake Tanganyika, C – Lake Malawi. Graphic image courtesy of esri.com (reproduced with permission).

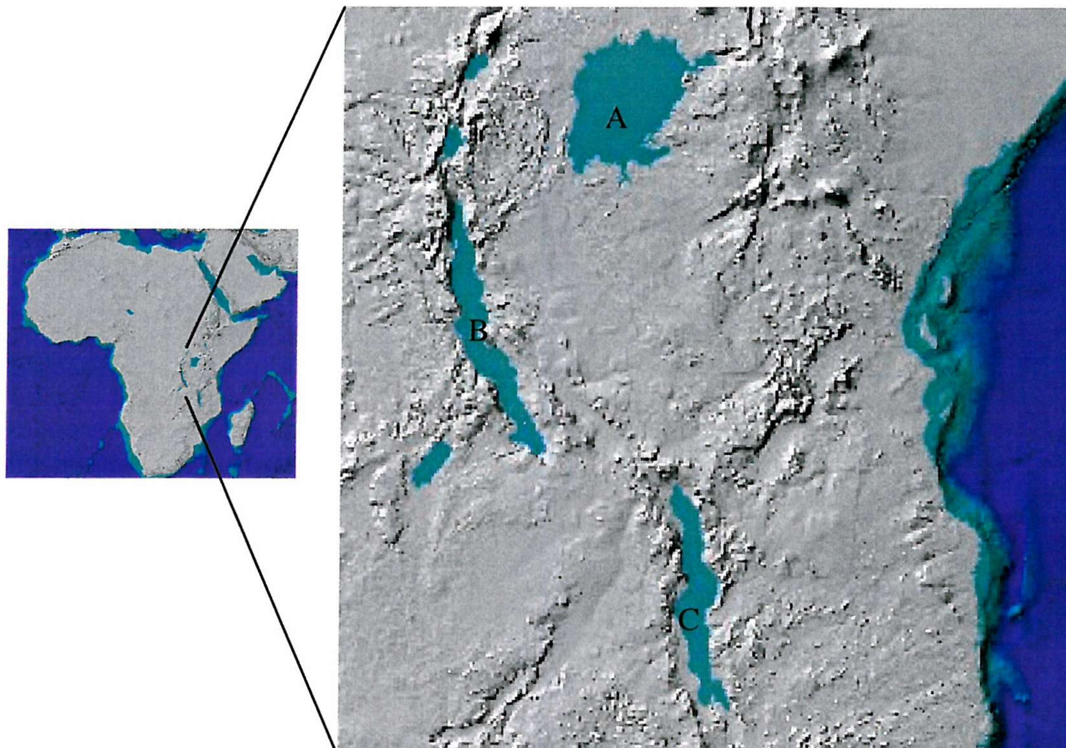


Table 1.1. Comparison of the physical size and magnitude of the cichlid radiation for the three Great Lakes. Age, depth and surface area data from Martens (1997), number of cichlid species from Turner (1999), age of radiation from Meyer (1993)

Lake	Age (MY)	Max Depth (m)	Surface Area (km^2)	Number of cichlid species	Age of Radiation (MY)
Victoria	0.5-1	80	70,000	250-500	0.2
Malawi	4-9	800	30,800	600-1000	0.7-1
Tanganyika	9-12	1470	32,600	170-200	4-9

Ribbink *et al* (1983) to classify populations that differed in male colour pattern as different species. Whilst the study of Ribbink *et al* (1983) did not provide formal descriptions of these newly classified species, it introduced a system of nomenclature that combined formal taxonomic names with anglicized colour and location descriptors. The ecology of the mbuna, together with the ease of observation and sampling by SCUBA, have resulted in mbuna species being used to test many hypotheses regarding evolutionary mechanisms in Lake Malawi. It should be noted that similar groups of cichlids also occupy the rocky littoral habitat of Lake Victoria and Tanganyika and have also been the object of evolutionary studies, although currently fewer in number.

The existence of such strong geographical population structuring has led to the suggestion that speciation in allopatry was, and is, an important force in the explosive speciation of Malawi cichlids. Genetic drift and differing selective pressures between habitat patches could result in reproductive isolation, so that even if dispersal became possible between the two populations, interbreeding would be prevented. Genetic population subdivision of four species of mbuna was measured directly by van Oppen *et al* (1997). It was found that populations separated by 700-1400m were genetically distinct from one another, which indicated restricted migration. This unusually fine scale structuring, if present throughout the lake, provides an enormous opportunity for allopatric speciation. Similar studies on other mbuna species in different areas of the lake also found significant genetic differences between populations, although there appeared to be no barriers to gene flow within habitat patches (Arnegard *et al*, 1999; Danley *et al*, 2001, Markert *et al*, 1999).

The question of what constitutes an mbuna species is still open to debate. McKaye *et al* (1984) demonstrated significant differences in electromorph (allozyme) frequencies between different colour morphs of the mbuna species *Metriaclima zebra*. Using microsatellites, it has been demonstrated that otherwise morphologically similar colour morphs do not interbreed in the wild (van Oppen *et al*, 1998) and in the laboratory (Knight *et al*, 1998). Ribbink *et al* (1983) treated each colour variant as a separate species, although many have yet to be formally described as such. However, it is known that certain of the mbuna will hybridise, especially when provided with no choice under laboratory conditions (e.g. Crapon de Caprona and Fritzsch, 1984; Albertson and Kocher, 2001; Knight *et al*, 1998). It is possible for populations to

diverge in male colour but not sufficiently to become reproductively isolated as Arnegard *et al* (1999) suggest in their study of the mbuna species *Labeotropheus fuelliborni* and as also suggested by Magurran (2000) in a review of research on Trinidadian guppies. An important isolating mechanism among the mbuna appears to be sexual selection – female preference for certain male traits e.g. colour patterns or long trailing fins. Knight *et al* (1998) demonstrated that under laboratory conditions, females of three mbuna species *Metriaclima zebra*, *M. zebra* species 'gold' and *M. callainos* mate assortatively. Assortative mating was determined by paternity assessment of offspring using microsatellites. In contrast, males of some mbuna species were unable to distinguish between conspecific and heterospecific females using vision alone (Knight and Turner, 1999) suggesting that female discrimination may be an important factor in reproductive isolation. Another study by Taylor *et al* (1998) indicated that female *Copadichromis eucinostomus*, a non-mbuna sand bower building species complex, choose males with reduced parasite loads and there was a correlation between parasite load and bower asymmetry. Studies such as these demonstrate the potential for sexual selection, which has been put forward as an isolating mechanism during sympatric speciation. Sexual selection has been postulated as causing divergence through random co-evolution of female preferences for male traits (e.g. Lande, 1981) and this may explain why populations diverge rapidly in male colour but not in other traits (Deutsch, 1997).

The occurrence of sympatric speciation remains controversial (see Futuyama, 1986), but is becoming more widely accepted (Via, 2001). Schliewen *et al* (1994 and 2001) reported strong evidence for the occurrence of sympatric speciation in West African crater lakes based on correlations of genetic and ecological divergence. A similar study on a Neotropical cichlid species found in Nicaraguan lakes by Wilson *et al* (2000) found evidence for sympatric genetic divergence with colour polymorphism. Sympatric speciation was invoked as an explanation of diversity within a pelagic cichlid genus from Lake Malawi by Shaw *et al* (2000) because of a lack of physical divisions within the habitat. Simulation modelling has allowed investigation of the biological conditions under which sympatric speciation may occur. The model of Turner and Burrows (1995) found sympatric speciation could occur due to reversal of female preference within small populations under particular selection regimes. Other models have considered the conditions necessary for reproductive isolation due to the

genetic architecture of loci coding for both reproductive and ecological characters (Dieckmann and Doebeli, 1999; Kondrashov and Kondrashov, 1999). Seehausen *et al* (1999) also provide evidence for the occurrence of both sympatric and allopatric speciation by examining male breeding colour polymorphisms in closely related cichlids from Lake Victoria, and investigating the numbers of polymorphisms that occurred in sympatric or allopatric species. Although mbuna are thought to be poor dispersers, genetic studies have revealed low-level migration between populations (for example van Oppen *et al*, 1997). If migration of females with different mate preferences (for example colour pattern) occurs between populations, then divergence between them might be arrested. On the other hand, males dispersing into populations with different female preferences would have little reproductive success as they are competing with local males that display more desirable coloration. Knight *et al* (1999) investigated dispersal on a local scale in two mbuna species, and found that males appeared to disperse further than females. This pattern of male biased dispersal would allow disruptive sexual selection, such as that modelled by Turner and Burrows (1995).

Other, more general models of the evolution of cichlids in Lake Malawi have been proposed. Sturmbauer (1998) described a multi-stage model, with different mechanisms acting to cause divergence at different times. This model was based on observations of similar groups (ecotypes) of cichlids in several of the lakes in the Rift Valley. The first stage consists of sympatric speciation followed shortly after by rapid diversification in ecology and morphology. The second stage represents preliminary allopatric diversification due to a lack of dispersal ability. The third stage (secondary allopatric diversification) consists of cycles of population contact and re-isolation (possibly due to lake level fluctuations), causing hybridisation and ecological niche shifts due to competition. The fourth stage describes further allopatric divergence and an approach to evolutionary stasis. A similar model based on the mbuna but applicable to flocks within other lakes was suggested by Danley and Kocher (2001) in which cladogenesis occurred in three main stages. The first stage was characterised by ecological diversification and resulted in two major clades, the sand dwellers and the mbuna. The second stage consisted mainly of trophic diversification and is proposed to have resulted in the groups now classified as mbuna genera (these genera are currently classified according to tooth and body morphology). Danley and Kocher (2001) proposed that the final stage of cladogenesis is driven by divergent sexual selection on

male breeding colour, as within each genus species are distinguishable on the basis of male colour pattern.

1.4.2 Investigating the evolutionary history of cichlids

In order to determine the mechanisms responsible for generating such a wealth of species, it is first necessary to reconstruct the evolutionary history of the taxa within each lake. While plausible global and regional (Figure 1.4) cichlid phylogenies have been estimated, the rapidity of intra-lake divergence together with instances of convergent evolution mean that suitable molecular or morphometric markers have been lacking (Kornfield and Smith, 2000). This has hindered work on determining critical evolutionary mechanisms. Various molecular markers have been used to reconstruct the phylogeny of the cichlids of Eastern Africa, and that of the Malawi radiation. A phylogeny of the family Cichlidae, including groups from India, Africa and South America was produced by Streelman *et al* (1998) who used a single copy nuclear marker as well as a microsatellite locus.

The evolutionary relationships between African cichlids, including lacustrine and riverine species from Eastern and Western Africa were at least partially revealed by Mayer *et al* (1998, Figure 1.4). The data for the phylogeny were obtained by sequencing an anonymous nuclear marker, and also the control and cytochrome b region of mitochondrial DNA. It was shown that the Tanganyikan flock is well differentiated from the Victoria and Malawi flocks, and the notion of the monophyletic origin of the Lake Victoria regional superflock was also well supported. However, it was not immediately apparent whether the Malawi flock is part of the Victoria superflock or evolutionarily separate. In a much-cited piece of research, Meyer *et al* (1990, see also Meyer, 1993) investigated evolutionary relationships between the three species flocks. Monophyly of all three flocks was determined using several mtDNA genes (Figure 1.3). Meyer *et al* discounted a polyphyletic origin of the Tanganyikan species flock because Lake Tanganyika appears to be older than the main lineages of cichlids within it. However finding a molecular marker that provides enough resolution to characterise all levels of intra-lake relationships has proved a problem; Meyer (1993) notes that no variation was present in 363bp of the cytochrome b gene and only 2-3 substitutions are present in 440bp of the mitochondrial control region of cichlids from

Lake Victoria. Other studies used RFLP (Moran and Kornfield, 1993 and 1995) and sequencing (Bowers *et al*, 1994) of mitochondrial DNA for analysis of phylogenetic and population divergence within the mbuna clade of Lake Malawi. Assessment of phylogenetic relationships among mbuna species were confounded by incomplete lineage sorting and the presence of shared ancestral polymorphisms by Moran and Kornfield (1993) who suggested that multiple independent marker loci would be required for this purpose. Recently the development of a new class of molecular markers (AFLPs, see Chapter 6) allowed Albertson *et al* (1999) to determine the evolutionary history of certain members of the mbuna clade of Lake Malawi (Figure 1.5). This technique allows genetic variation to be sampled from many points across the genome and bypasses the problems of low variability in single genes. The authors compared jaw morphology to the resulting groupings, and found that similar morphologies were clustered. However, the genera included in this study were separated by only extremely short branch lengths, this indicating extremely rapid divergence or a lack of phylogenetic resolution. A lack of resolution was discounted as both deeper and more recent divergences were recovered from the data. Nonetheless, AFLPs seem be a suitable marker for studying intra-lake and even intra-clade divergences, and it is possible that evolutionary relationships between members of the closely related mbuna group may be resolved using this technique.

1.4.3 A note on mbuna taxonomy

The taxonomy of part of the genus *Pseudotropheus* was revised by Stauffer *et al* (1997). The species complex *P. (Maylandia)* was promoted to genus status and renamed *Metriaclima*. This has caused controversy as some authorities feel that *Metriaclima* is a junior synonym of *Maylandia* (i.e. the generic name should be *Maylandia*). In addition, some authors now refer to *P. (Maylandia)* as *P. (Metriaclima)*. This thesis follows the nomenclature of Stauffer *et al* (1997) by using *Metriaclima* as a generic name. Members of the species complex referred to as *P. (Tropheops)* are referred to under the generic name of *Tropheops* following Konings (2001).

Figure 1.3. Phylogenetic tree based on mtDNA data redrawn from Meyer (1993). Shaded boxes are presumed monophyletic clades. The main clades of Lake Malawi are indicated in grey.

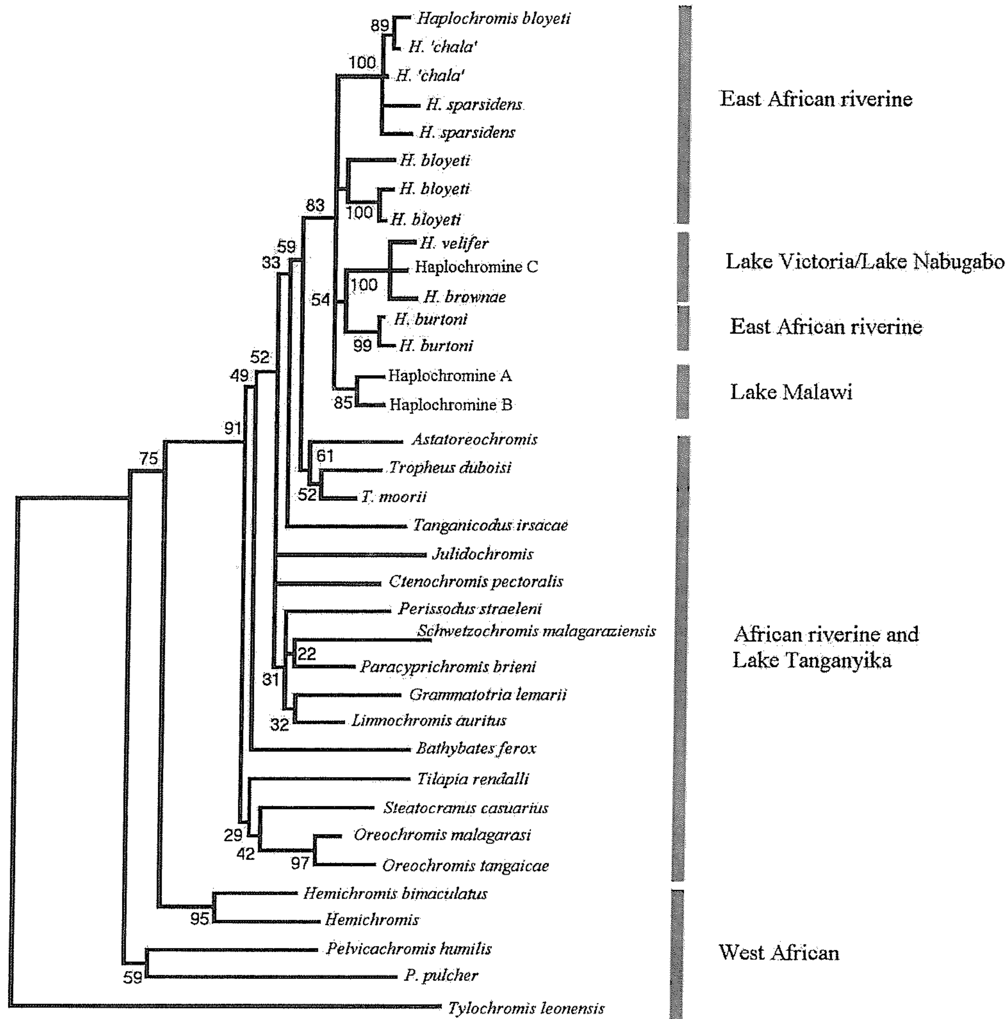


Figure 1.4. Phylogenetic relationships of African cichlids based on cytochrome b sequences. It can be seen that some nodes are not resolved, and others have low levels of bootstrap support. Haplochromine A and B represents the sand dwellers and the mbuna respectively and Haplochromine C represents a clade of species from Lake Victoria. Redrawn from Mayer *et al* (1998) with permission.

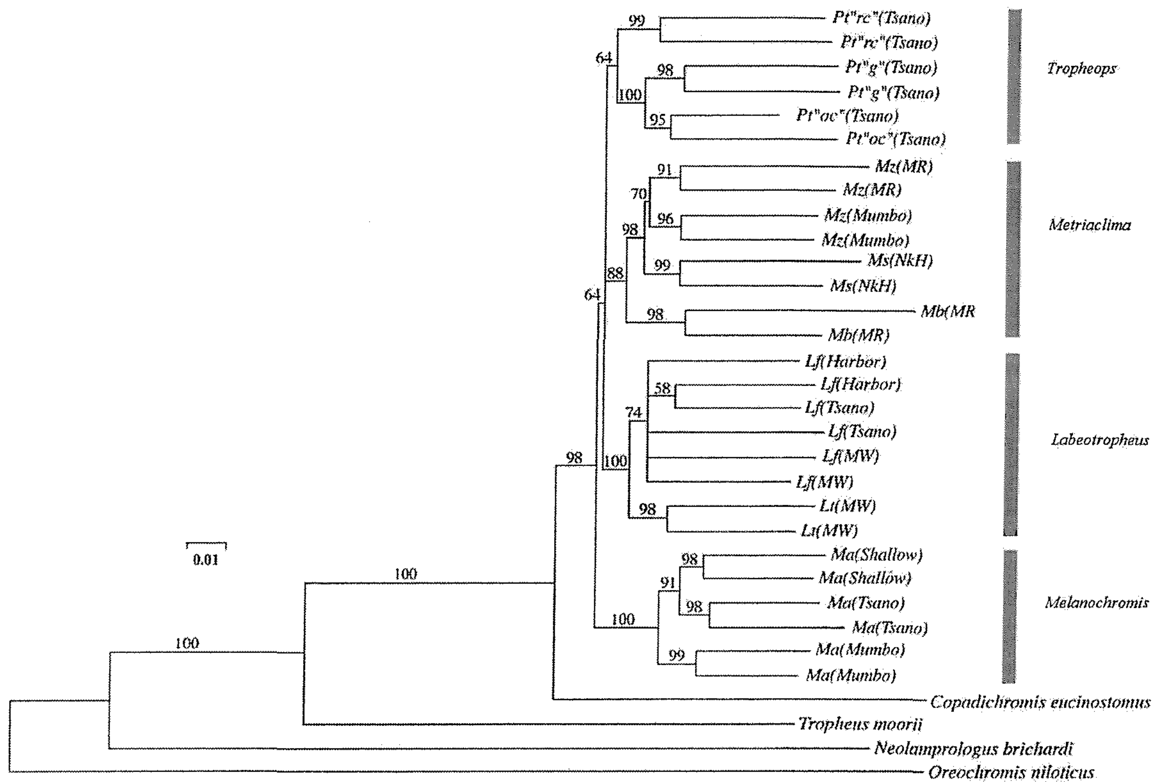


Figure 1.5. The phylogenetic reconstruction of four mbuna genera produced using AFLPs by Albertson *et al* (1999). Numbers at nodes represent the number of time the node was recovered in 100 bootstrap replications. The scale bar indicates 1% divergence in AFLP profile. Populations cluster into genera with high bootstrap support. Reproduced with permission, original figure © 1999 National Academy of Sciences of the USA.

1.4.4 Summary

It is probable that a number of speciation mechanisms have operated in Lakes Malawi, Tanganyika and Victoria, and indeed may still be causing divergence of populations. Microsatellites in particular have proved to be very useful in ascertaining mating systems and dispersal patterns. Together with conventional field and laboratory techniques, as well as newer molecular methods such as AFLP, microsatellites should be able to add further detail to the picture of cichlid evolution in the African Great Lakes.

1.5 Aims and Scope of Thesis

This study aims to further the use of molecular markers in the study of speciation in African cichlid fish. As discussed above, female mate choice may be an important factor in initiating and maintaining divergence between populations. Highly polymorphic microsatellite DNA has proved to be a useful marker to use in studies of mate choice, and is applied here to two investigations testing whether females can discriminate between differently coloured males belonging to sympatric and allopatric populations. Polymorphic microsatellites can also be used to provide a statistical estimate of the level of relatedness between two individuals (this has already been used to demonstrate the occurrence of male biased dispersal in an mbuna species by Knight *et al*, 1999), and the number of loci required for the calculation of a relatedness statistic is assessed. Robust phylogenies of the mbuna clade have proved difficult to obtain, and this has prevented hypothesis testing regarding evolutionary patterns in this group. This problem is addressed using two different classes of molecular markers, microsatellites and AFLPs.

1.6 References

Albertson RC and Kocher TD (2001) Assessing morphological differences in an adaptive trait; a landmark based morphometric approach. *Journal of Experimental Zoology* 289: 385-403

Albertson RC, Markert JA, Danley PD and Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of. Sciences of the USA* 96 5107-5110

Allender CJ (1998) Using Randomly Amplified Polymorphic DNA to Study Genetic Diversity within Populations of the Hazel Dormouse *Muscardinus avellanarius*. MRes Dissertation, University of Warwick, UK.

Angers B and Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution* 14: 230-238

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ and Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelliborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London B* 266: 119-130

Avise JC (1994) Molecular Markers, Natural History and Evolution. Chapman and Hall, New York

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA and Saunders NC (1987) Intraspecific phylogeny; the mitochondrial bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*. 18: 498-522

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

Booton GC, Kaufman L, Chandler M, Ogto-Ohwayo R, Duan W and Fuerst PA (1999) Evolution of the Ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake Victoria Region. *Molecular Phylogenetics and Evolution* 11: 273-282

Bowers N, Stauffer JR and Kocher TD (1994) Intra- and interspecific mitochondrial DNA sequence variation within two species of rock-dwelling cichlids within Lake Malawi, Africa. *Molecular Phylogenetics and Evolution* 3: 75-82

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. *Methods in Enzymology* 29: 363-418

Brownstein MJ, CarptenJD and Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20: 1004-1010

Cavalli-Sforza LL and Edwards AWF (1967) Phylogenetic analyses: models and estimation procedures. *Evolution* 32: 550-570

Crapon de Caprona M and Fritzsch B (1984) .... *Netherlands Journal of Zoology* 34: 503-538
Danley PD and Kocher TD (2001) Speciation in Rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology.....*

Danley PD, Markert JA, Arnegard ME and Kocher TD (2001) Divergence with gene flow in the rock dwelling cichlids of Lake Malawi. *Evolution*:

Deutsch JC (1997) Colour diversification in Malawi cichlids: evidence for adaptation, reinforcement or sexual selection. *Biological Journal of the Linnean Society* 62: 1-14

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M and Friemer NB (1994) Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA* 91: 3166-3170

Dieckmann U and Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* 400: 354-357

Estoup A and Angers B (1998) Microsatellites and minisatellites for molecular ecology : theoretical and empirical considerations. In *Advances in Molecular Ecology* ed GR Carvalho. IOS Press.

Feldman MW, Bergman A, Pollock DD, and Golstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* 145: 207-216

Fryer G (2001) On the age and origin of the haplochromine cichlid fishes of Lake Victoria. *Proceedings of the Royal Society of London B* 268: 1147-1152

Fryer GT and Iles T (1972) The cichlid fishes of the great lakes of Africa. Oliver and Boyd, London.

Futuyama DJ (1986) Evolutionary Biology. Sinauer Inc, Sunderland

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

Goldstein DB and Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 88: 335-342

Goldstein DB, Linares AR, Cavalli-Sforza LL and Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL and Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proccedings of the National Academy of Sciences of the USA* 92 : 6723-6727

Grant PR (1993) Hybridisation of Darwin's finches on Isla Daphne Major, Galapagos. *Philosophical Transactions of the Royal Society of London B* 340: 127-139

Greenwood PH (1965) The cichlid fishes of Lake Nabugabo, Uganda. *Bulletin of the British Museum of Natural History (Zoology)* 12: 315-357

Greenwood PH (1984) African cichlid fishes and evolutionary theories. In *Evolution of Fish Species Flocks* (eds Echelle AA and Kornfield I) pp141-154

Hadrys H, Schierwater B, Dellaporta SL, DeSalle R and Buss RW (1993) Determination of paternity in dragonflies by Random Amplified Polymorphic DNA fingerprinting. *Molecular Ecology* 2: 79-87

Hamby RK and Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In *Molecular Systematics of Plants* eds PS Soltis, DE Soltis, JJ Doyle. Chapman and Hall, New York

Harris H (1966) Enzyme polymorphisms in man. *Proceedings of the Royal Society of London B* 164: 298-390

Hillis DM, Mable BK, Larson A, Davis SK and Zimmer EA (1996) Nucleic acids IV: sequencing and cloning. In *Molecular Systematics* 2nd edn eds DM Hillis, C. Moritz and BK Mable. Sinauer

Hunt DM, Fitzgibbon J, Slobodyanyuk SJ, Bowmaker JK and Dulai KS (1997) Molecular evolution of the Cottoid fish endemic to Lake Baikal deduced from nuclear DNA evidence. *Molecular Phylogenetics and Evolution* 8: 415-422

Jacobsen R, Forbes VE and Skovgaard O (1996) Genetic poulation structure of the prosobranch snail Potomopyrgus antipodum in Denmark using PCR-RAPD fingerprints. *Proceedings of the Royal Society of London B* 263: 1065-1070

Jarne P and Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424-429

Jefferys AJ, Royle NJ, Wilson V and Wong Z (1988) Spontaneous mutations rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332: 278-281

Jefferys AJ, Wilson V and Thein SL (1985) Hypervariable 'minisatellite' regios in human DNA. *Nature* 314: 67-73

Johnson FM, Kanapi CG, Richardson RH, Wheeler MR and Stone WS (1966) An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from America and Western Samoa. *Proceedings of the National Academy of Sciences of the USA* 56: 119-125

Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD, Ngobi G, Beuning K, Ssemunda I and McGill JW (1996) Late Pleistocene dessication of Lake Victoria and rapid evolution of cichlid fishes. *Science* 273: 1091-1093

Jones AG, Stockwell CA, Walker D and Avise JC (1998) The molecular basis of a microsatellite null allelle from the white sands pupfish. *Journal of Heredity* 89: 339-342

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Van de Wiel C, Bredermeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volkaert G, Rueda J, Linacero R, Vaques A and Karp A (1997) Reproducability testing of RAPD, AFLP and SSCP markers in plants by a network of European laboratories. *Molecular Breeding* 3: 381-390

Kimura M and Crow JF (1964) The number of alleles that can be maintained in a population. *Genetics* 49: 725-738

Knight ME and Turner GF (1999) Reproductive isolation in closely related Lake Malawi

cichlids: can males recognise females by visual cues? *Animal Behaviour* 58: 761-768

Knight ME, Turner GF, Rico C, van Oppen MJH and Hewitt GM (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology*: 1605-1610

Knight ME, van Oppen MJH, Smith HL, Rico C, Hewitt GM and Turner GF (1999) Evidence for male biased dispersal in Lake Malawi cichlids from microsatellites. *Molecular Ecology* 8: 1521-1527

Kocher TD, Lee WJ, Soboloweska H, Penman D and McAndrew B (1998) A genetic linkage map of a cichlid fish, the Tilapia (*Oreochromis niloticus*). *Genetics* 148: 1225-1232

Kondrashov AS and Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400: 351-354

Konings A (2001) *Malawi Cichlids in their Natural Habitat.* 3rd Edn. Cichlid Press

Kornfield I and Smith PF (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Review of Ecology and Systematics* 31: 163-196

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research* 21: (5) 1111-1115

Magnuson VL, Ally DS, Nylund SJ, Karajawala ZE, Rayman JB, Knapp JI, Lowe AL, Ghosh S and Collins FS (1996) Substrate nucleotide determined non templated addition of adeneine by Taq DNA polymerase : Implications for PCR based genotyping and cloning. *Biotechniques* 21: 700-709

Magurran AE (1998) Population differentiation without speciation. *Philosphical Transactions of the Royal Society Series B* 353: 275-286

Markert JA, Arnegard ME, Danley D and Kocher TD (1999) Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*; habitat transience, philopatry and speciation. *Molecular Ecology* 8: 1013-1026

Martens K (1997) Speciation in ancient lakes. *Trends in Ecology and Evolution* 12: 177-182

Mayer WE, Tichy H and Klein J (1998) Phylogeny of African cichlid fishes as revealed by molecular markers. *Heredity* 80: 702-714

McEwan NR, Bahkt J, Cecchini E, Coulhart A, Geri C, McDonald FM, McDonald MS, Rahman NNA and Williamson G (1998) Examination of the pipetting variation between individuals performing random amplified polymorphic DNA (RAPD) analysis. *Journal of Microbiological Methods* 32: 213-215

McKaye KR, Kocher T, Reinthal P, Harrison R and Kornfield I (1984) Genetic evidence for allopatric and sympatric differentiation among color morphs of a Lake Malawi cichlid fish. *Evolution* 38: 215-219

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution* 8: 279-284

Meyer A, Kocher TD, Basasibwaki P and Wilson AC (1990) Monophyletic Origin of Lake

Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347: 550-553

Moran P and Kornfield I (1993) Retention of an ancestral polymorphism in the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution* 10: 1015-1029

Moran P and Kornfield I (1995) Were population bottlenecks associated with the radiation of the mbuna secies flock (Teleostei: Cichlidae) of Lake Malawi? *Molecular Biology and Evolution* 12: 1085-1093

Mueller UG and Wolfenbarher LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* 14: 389394

Mullis KB and Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. *Methods in Enzymology* 155: 335-350

Nagl S, Tichy H, Mayer W, Takeszaki N, Takhhata N and Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proceedings of the Royal Society of London B* 267:1049-1061

Nei M, Tajima F and Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19: 153-170

Ohta and Kimura (1973) The model of mutation appropriate to estimate the number of electrophetically detectable alleles in a genetic population. *Genetic Research* 22: 201-204

Olsen JB, Wenburg JK and Bentzen P (1996) Semiautomated multilocus genotyping of Pacific salmon (*Oncorhnychus* spp.) using microsatellites. *Molecular Marine Biology and Biotechnology* 5: 259-272

Orti G, Pearse DE and Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences USA* 94: 10745-10749

Otte D (1989) Speciation in Hawaiian crickets. In *Speciation and its Consequences* eds D Otte and JA Endler. Sinauer, Sunderland MA

Owen BR, Crossley R, Johnson TC, Tweddle D, Kornfield I, Davison S, Eccles DH and Engstrom DE (1990) Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proc. R. Soc. Lond. B* 240: 519-553

Paektau D, Waits LP, Clarkson PL, Craighead L and Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147: 1943-1957

Parker PG, Snow AA, Schug MD, Booton GC and Fuerst PA (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* 79: 361-382

Parsons YM and Shaw KL (2001) Species boundaries and gentic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology* 10: 1765-1772

Petren K, Grant BR and Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proceedings of the Royal Society of London B* 266 : 321-329

Powell JR and Desalle R (1995) *Drosophila* molecular phylogenies and their uses.

Evolutionary Biology. 28: 87-138

Quagliaro G, Vischi M, Tyrka M and Olivieri AM (2001) Identificatiuon of wild and cultivated sunflower for breeding purposes by AFLP markers. *Genetical Research* 92: 38-42

Ribbink AJ (1994) Lake Malawi. *Arch. Hydrobiol. Beih. Ergebn. Limnol* 44 27-33. Speciation in Ancient Lakes (ed K. Martens, B. Goddeeris and G. Coulter)

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC and Sharp BJ (1983) A preliminary survey of the cichlid fishes of rocky habitats of Lake Malawi. *S. Afr. J. Sci.* 18: 149-310

Rico C, Rico I and Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London* B 263: 549-557

Saika RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB and Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491

Savelkoul PHM, Aarts HJM, de Haas J, Dijkshoorn L, Duin B, Otsen M, Rademaker JLW, Schouls L and Lenstra JA (1999) Amplified-fragment length polymorphism analysis: the state of an art. *Journal of Clinical Microbiology* 37: 3083-3091

Schliewen UK, Tautz D and Paabo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368: 629-632

Schlotterer (1998) Microsatellites. In *Molecular Genetic Analysis of Populations – a Practical Approach*, ed AR Hoelzel IRL Press

Schlotterer C and Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20: 211-215

Schliewen U, Rassman K, Markmann M, Markert J, Kocher T and Tautz D (2001) Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology* 10: 1471-1488

Sebastian RL, Howell EC, King GJ, Marshall DF and Kearsey MJ (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid populations. *Theoretical and Applied Genetics* 100: 75-81

Seehausen O, van Alpen JJM (1998) The effect of male colouration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioural Ecology and Sociobiology* 42: 1-8

Sharbel TM, Haubold B and Mitchell-olds T (2001) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonisation of Europe. *Molecular Ecology* 9: 2109-2118

Shaw KL (1996) Sequential radiations and patterns of speciation in the Hawaiian cricket genus *Laupala* inferred from DNA sequences. *Evolution*: 237-255

Shaw PW, Turner GF, Idid MR, Robinson RL and Carvaho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society of London* B 267: 2273-2280

Shriver MD, Jin L, Chakraborty R and Boerwinkle E (1993) VNTR allele frequency

distributions under the stepwise mutation model: a computer simulation approach. *Genetics* 139: 983-993

Sibley CG and Ahlquist JE (1990) *Phylogeny and classification of birds*. Yale University Press, New Haven CT

StaufferJR, Bowers NJ, Kellogg KA and McKaye KR (1997) A revision of the blue-black *Pseudotropheus zebra* (Teleostei:Cichlidae) complex from Lake Malawi, Africa, with a description of a new genus and ten new species. *Proceedings of the Academy of Natural Sciences of Philadelphia* 148: 189-230

Sideleva VG (2000) The ichthyofauna of Lake Baikal, with special reference to its zoogeographical relations. In *Advances in Ecological Research* Special Issue: Biology of Ancient Lakes (eds Rossiter A, Kawanabe H) pp 81-96.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462

Stephens JC, Gilbert DA, Yuhki N and O'Brien SJ (1992) Estimation of heterozygosity for single-probe multilocus DNA fingerprints. Molecular Biology and *Evolution* 9: 729-743

Streelman JT, Zardoya R, Meyer A and Karl SA (1998) Multilocus phylogeny of cichlid fishes (Pisces : Perciformes) : evolutionary comparison of microsatellite and single copy nuclear loci. *Molecular Biology and Evolution* 15: 798-808

Sturmbauer C (1998) Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. *Journal of Fish Biology* 53 (SA) 18-36

Sturmbauer C, Baric S, Salzburger W, Ruber L and Verheyen E (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African Lakes. *Molecular Biology and Evolution* 18: 144-154

Surridge AK, Ibrahim KM, Bell DJ, Webb NJ, Rico C, and Hewitt GM (1999) Fine-scale genetic structuring in a natural population of European wild rabbits (*Oryctolagus cuniculus*). *Molecular Ecology* 8: 299-307

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471

Taylor AC, Horsup A, Johnson CN, Sunnucks P and Sherwin B (1997) Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy nosed wombat *Lasorhinus krefftii*. *Molecular Ecology* 6: 9-19

Taylor JS, Durkin, JMH and Breden F (1999) The death of a microsatellite: a perspective on microsatellite interruptions. *Molecular Biology and Evolution* 16 : 567-572

Taylor MI, Turner GF, Robinson R and Stauffer JR (1998) Sexual selection, parasites and bower height skew in a bower building cichlid fish. *Animal Behaviour* 56: 379-384

Turner GF (1999) Explosive Speciation of African Cichlid Fishes. In *Evolution of Biological Diversity* eds AE Magurran and RM May Oxford University Press

Turner GF and Burrows MT (1995) A model of sympatric speciation by sexual selection. *Proceedings of the Royal Society of London Series B* 260: 287-292

Valdes am, Slatkin M and Freimar NB (1993) Allele frequencies at microsatellite loci revisited. *Genetics* 133: 737-749

van Oppen MJH, Turner GF, Rico C, Deutsch JC, Ibrahim KM, Robinson RL and Hewitt GM (1997) Unusually fine scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London Series B* 264: 1803-1812.

van Oppen MJH, Turner GF, Rico C, Robinson RL, Deutsch JC, Genner MJ and Hewitt GM (1998) Assortative mating among rock dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Molecular Ecology* 7: 991-1001

Via S (2001) Sympatric speciation in animals; the ugly duckling grows up. *Trends in Ecology and Evolution* 16: 381-389

Vincek V, O'Huigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR and Klein K (1997) How large was the pounding population of Darwin's finches? *Proceedings of the Royal Society of London B* 264: 111-118

Vos P, Hogers R, Bleeker M (1995) AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Watt WB, Cassin RC and Swan MS (1983) Adaptation at specific loci, III. Field behaviour and survivorship differences among *Colias* PGI genotypes are predictable from in vivo biochemistry. *Genetics* 103: 725-739

Wattier R, Engel CR, Saumitou-Laprade P and Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucletide locus Gv1CT in *Gracillaria gracilis* (Rhodophyta). *Molecular Ecology* 7: 1569-1573

Webb NJ, Ibrahim KM, Bell DJ and Hewitt GM (1995) Natal dispersal and genetic structure in a population of the European wild rabbit. *Molecular Ecology* 4: 239-247

Weber JL and Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123-1128

Weiss RE, Carmack EC and Koropalov VH (1991) Deep water renewal and biologica productivity in Lake Baikal. *Nature* 349: 665-669

Wenburg JK, Bentzen P and Foote CJ (1998) Microsatellite analysis of genetic population structure in an endangered salmonid : the coastal cutthroat trout (*Oncorhynchus clarki clarki*). *Molecular Ecology* 7: 733-749

Wenburg JK, Olsen JB and Bentzen P (1996) Multiplexed systems of microsatellites for genetic analysis in coastal cutthroat trout (*Oncorhynchus clarki clarki*) and steelhead (*Oncorhynchus mykiss*). *Molecular Marine Biology and Biotechnology* 5: 273-283

Werman SD, Springer MS and Britten RJ (1996) Nucleic acids I: DNA-DNA hybridisation. In *Molecular Systematics* 2nd edn eds DM Hillis, C. Moritz and BK Mable. Sinauer

West K and Michel E (2000) The dynamics of endemic diversification: molecular phylogeny suggests an explosive origin of the Thiarid gastropods of Lake Tanganyika. In *Advances in Ecological Research* Special Issue: Biology of Ancient Lakes (eds Rossiter A, Kawanabe H) pp331-354.

Williams JGK, Kubelik Ar, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531-6535

Wilson AB, Noack-Kunnman K and Meyer A (2000) Incipient speciation in Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proceedings of the Royal Society of London B* 267: 2133-2141

Wright S (1951) The genetical structure of populations. *Ann Eugen.* 15:323-354

Zardoya R, Vollmer DM, Craddock C, Streelman GT, Karl S and Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their uses in resolving the phylogeny of cichlid fishes. *Proceedings of the Royal Society of London B* 263: 1589-1598

2 Improving Microsatellite Based Estimates of Relatedness

2.1 Introduction

Knowledge of the levels of relatedness between pairs or groups of individuals is important as it allows for the testing of hypotheses regarding ecological and behavioural parameters. Hamilton (1964) significantly clarified the importance of the relatedness of two interacting individuals, particularly when altruistic behaviour is involved. Relatedness can be calculated directly using pedigrees where parentage is known. However, this method is generally only applicable to captive or heavily monitored populations. Reliable pedigrees are difficult to obtain in previously unstudied populations and also when multiple matings occur. The advent of DNA markers has allowed relatedness to be assessed indirectly on the basis of genetic similarity. Microsatellites have proved particularly useful in this respect due to their co-dominant nature and generally high levels of polymorphism (Queller *et al*, 1993).

Relatedness estimates can be used to examine social and breeding structure among group-living animals. For example Smith *et al* (1997) used relatedness estimates to assess kinship among reproductive members of grey wolf (*Canis lupus*) packs, and hence to estimate potential levels of inbreeding. Reproductive success of male grey seals (*Halichoerus grypus*) was investigated by Worthington Wilmer *et al* (2000) who looked at kinship relationships of seal pups in a single breeding location. Relatedness estimates can also provide information on other aspects of the behavioural ecology of a species such as dispersal. One would expect to find clusters of individuals showing higher relatedness than the population average if those individuals do not generally disperse away from their natal area. If neither sex disperses then obviously the potential for inbreeding increases greatly. However, many species exhibit sex-biased dispersal, with either females or males more likely to move away from their natal area. Again, this can be demonstrated by finding higher than average clusters of relatedness in one sex but not in the other. This kind of dispersal has the obvious potential for the reduction of inbreeding within a population (Chesser, 1991). Sex-biased dispersal has been demonstrated through relatedness statistics in European wild rabbits (Surridge *et*

al, 1999), the northern hairy nosed wombat (Taylor *et al*, 1997), red grouse (Piertney *et al*, 1999) and Lake Malawi cichlids (Knight *et al*, 1999).

The conclusions of studies based on statistics of relatedness rely on obtaining valid estimates. However, relatedness estimators are often associated with a high level of variance. It is acknowledged that one of the methods of reducing this variance is to use a higher number of loci in calculations (van de Casteele *et al*, 2001; Lynch and Ritland, 1999). Brookfield and Parkin (1993) suggested that an unfeasibly large number of unlinked loci would be needed to accurately discriminate between anything other than first order relatives (for example cousins and unrelated individuals). However, many questions involving relatedness do not require such discrimination. Blouin *et al* (1996) were able to distinguish between pairs of full siblings and unrelated individuals with 97% accuracy, based on 20 microsatellite loci. They also discovered that the number and heterozygosity of the loci used had a noticeable effect on the number of misclassifications; more loci of lower variability were required to achieve the same level of discrimination.

A brief survey of microsatellite based relatedness studies revealed variability in the numbers of loci employed (Table 2.1). Total numbers of loci used ranged from four to twenty. This probably reflects both the nature of the problem addressed and the number and type of loci available for the organism of interest. However, it is often not possible or desirable to type large numbers of individuals at many loci. Indeed, it would be extremely useful to know the minimum number of loci required for investigations of relatedness in advance, in order to minimise the cost and time associated with molecular analyses. An unequivocal method of testing the accuracy of relatedness statistics is to calculate them for known relatives. This study aims to use a set of known relatives to determine the effect of employing more loci on Queller and Goodnight's (1989) estimate of Hamilton's R. This estimator is currently the most commonly used for studies of relatedness with 193 citations since 1996 (Web of Science Citation Database search). In a wider context, this study also aimed to test and optimise five extra loci for future investigations on Malawi cichlids.

Table 2.1. Numbers of loci employed by recently published microsatellite based relatedness studies. The mean number of loci employed ranges from 10.9 to 12.0 as some studies utilised varying numbers of loci.

Authors	Species	No. Loci used
Surridge *et al* (1999)	*Oryctolagus cuniculus*	10
Worthington Wilmer *et al* (2000)	*Halichoerus grypus*	9
Piertney *et al* (1999)	*Lagopus lagopus scoticus*	12-17
Smith *et al* (1997)	*Canis lupus*	20
Taylor *et al* (1997)	*Lasiorhinus krefftii*	8-9
Favre *et al* (1997)	*Crocidura russala*	4 or 8
Blouin *et al* (1996)	*Mus musculus*	20
Kays *et al* (2000)	*Potos flavus*	11
Pouyard *et al* (1999)	*Sarotherodon melanotheron*	4

2.2 Methodology

2.2.1 Sample Origin and DNA Extraction

Three broods of two species of Malawi cichlids were available for typing, one of *Metriaclima callainos*, and two of the closely related, sympatric *M. zebra*. The same male sired both *M. zebra* broods, but the mothers were unrelated. Genetic samples from adults were obtained in the form of clips from the soft-rayed posterior portion of the dorsal fin. DNA from the offspring was extracted from whole eggs or fin clips from fry. All genetic samples were preserved in DMSO buffer (Seutin *et al*, 1991) prior to extraction. DNA was extracted by a proteinase K/chloroform protocol (Rico *et al*, 1992). Briefly, the entire egg/larva or half the fin clip was washed in Millipore water and placed in 600 µl SE buffer (2% SDS, 25mM EDTA, 75 mM NaCl). 20 µl 10mg/ml proteinase K was added and the mixture was incubated for 2 hours at 50-53°C in a water bath. After incubation, 290 µl 5M NaCl and 1 ml chloroform was added and the samples were thoroughly mixed for 15 minutes on a rotor. Samples were then centrifuged at 12000 rpm for 15 minutes using an Eppendorf 5102 benchtop microcentrifuge. The aqueous phase was removed and place in a clean tube and 1 ml of

ice-cold isopropanol was added. Samples were then left overnight in a –20°C freezer. The DNA was then pelleted by centrifugation for 15 minutes at 12000 rpm. The resulting pellets were washed twice by ice-cold 70% ethanol and left to air dry. Pellets were then resuspended in 200 µl (fin clips and larvae) or 100 µl (eggs) Tris-HCl buffer (10 µM, pH 9.0).

2.2.2 *PCR Conditions*

The 11 loci chosen were polymorphic dinucleotide repeats (Table 2.2). Not all loci were originally isolated from the species under investigation. Two were originally isolated from *Oreochromis niloticus* (UNH106 and UNH130), eight from *M. zebra* (Pzeb1-5, UNH002, UME003 and DXTUCA-3) and one from *Tropheus moorii* (TmoM11). Typing at the first six loci (Pzeb1-5 and UNH002) had already been carried out (Knight *et al*, 1998). PCR conditions were similar for all five of the remaining loci typed in this study. Reactions were carried out in 11 µL volumes containing approximately 9.9ng of each primer, 20µm each dNTP, 1.1µM magnesium chloride, 0.2 µg BSA and 0.05 units Thermoprime plus polymerase (Abgene). DNA extract was used undiluted or diluted 1 in 10. A Hybaid Touchdown thermal cycler was used to carry out the following: 5 minutes at 94°C followed by five cycles of 94°C (45s), 58°C (45s) and 72°C (30s), then 30 cycles of 91°C (30s), 56°C (30s) and 72°C (20s). For locus UNH106, annealing temperatures were reduced to 53 and 51°C respectively. Forward primers were labelled at 5^1 with FAM, TET or HEX to enable sizing by an ABI 377 (Applied Biosystems), a semi-automated sequencer. PCR products were diluted 1 in 5 or 1 in 10 depending on their relative concentrations as determined by agarose gel electrophoresis, and products tagged with different dye labels were mixed together for loading onto the sequencer. 0.5 µl of the diluted product mix was combined with 0.25 µl size standard (Genescan 350-TAMRA), 1.08 µl deionised formamide and 0.17 µl loading dye. This loading mix was then denatured at 90°C for 2 min and kept on ice until loading. Samples were run on a 36 cm well-to-read denaturing polyacrylamide gel (5.0 % GenePAGE Plus gel mix, 6 M urea) on an ABI 377 sequencer (Applied Biosystems) for 2 hours at the default electrophoresis parameters (3000V, 60mA, 200W at 51°C). After electrophoresis, fragments were sized and analysed using the GeneScan software package (Applied Biosystems).

2.2.3 Statistical Analysis

2.2.3.1 Linkage

The possibility of linkage between the 11 loci was assessed using the genotypic linkage disequilibrium test in the software package Genepop v3.3 (Raymond and Rousset, 1995). Markov chain parameters were set as 1000 dememorisation steps with 1000 batches of 10000 iterations per batch.

2.2.3.2 Pairwise Relatedness

Queller and Goodnight's (1989) R statistic (Equation 1) was calculated for each pair of individuals within the *M. zebra* broods (n = 18) and the *M. callainos* brood (n = 5) using Kinship 1.2 (see Goodnight and Queller, 1999). Only those individuals that amplified at all loci were included. Pairwise relatedness values were sorted into four categories depending on the pedigree relationship of the individuals involved (full sibs, parent/offspring, half sibs and unrelated individuals). Pairwise relatedness was calculated using different groups of loci with varying average heterozygosity (Table 2.2). The deviation of the R values from the expected theoretical values (0.5 for full sibs and parent/offspring pairs, 0.25 for half sibs and 0 for unrelated pairs) was also calculated to assess the accuracy of results using each group of loci. The range of R values generated for each relatedness category was used as an indication of the precision of the statistic.

Locus	Repeat Unit in Original Clone	Size Range	Number of Alleles	HE	Species of Origin	Authors
Pzeb1	$(GT)_{39}$	134-242	34	0.96	*Metriaclima zebra*	Van Oppen *et al* (1997a)
Pzeb2	$(GT)_{41}$	204-286	29	0.95	*Metriaclima zebra*	Van Oppen *et al* (1997a)
Pzeb3	$(GT)_{11}$	314-338	12	0.74	*Metriaclima zebra*	Van Oppen *et al* (1997a)
Pzeb4	$(GT)_4(TTGT)_2CTGCCT(GC)_{10}GC(GT)_2CT(GT)_5$ $GC(GT)_2$	123-145	10	0.77	*Metriaclima zebra*	Van Oppen *et al* (1997a)
Pzeb5	$(CT)_2TTT(CT)_8CA(CT)_2GTCT$	121-137	6	0.34	*Metriaclima zebra*	Van Oppen *et al* (1997a)
UNH002	$(CA)_{23}$	174-251	33	0.91	*Metriaclima zebra*	Kellogg *et al* (1996)
UNH106	$(CT)_{13}(CA)_{20}$	141-177	17	0.82	*Oreochromis niloticus*	Lee and Kocher (1998)
UNH130	$(CA)_{23}$	176-246	29	0.95	*Oreochromis niloticus*	Lee and Kocher (1998)
UME003	$(AT)_6AAA(AT)_2ACA(TG)_6(GCGT)_{13}$	193-245	24	0.93	*Metriaclima zebra*	Parker and Kornfield (1997)
DXTUCA-3	$(AC)_{37}$	103-161	18	0.93	*Metriaclima zebra*	Sueltmann *et al* (unpubl.)
M11	$(AC)_{19}$	203-233	19	0.94	*Tropheus moorii*	Zardoya *et al* (1996)

Table 2.2 Characteristics and origin of the 11 microsatellite loci used in this study. Data for loci Pzeb1-Pzeb5 are taken from van Oppen *et al* (1997b). Data for UNH002 taken from Kellogg *et al* (1996). Data for the remaining loci were obtained from 45 individuals of *M. callainos* sampled from Nkhata Bay.

Equation 1

$$\mathbf{R} = \frac{\sum_{x}\sum_{k}\sum_{l} (P_y - P^*)}{\sum_{x}\sum_{k}\sum_{l} (\mathbf{P_x - P^*})}$$

In this equation described by Queller and Goodnight (1989), *x* indexes individuals in the data set, *k* indexes loci and *l* indexes allelic position (*l*=1 or 2 for a diploid). P_x represents for individual x the frequency of the allele found at locus k and allelic position l (in a diploid either 0.5 or 1), P_y represents the frequency of the same allele in the individual which is being compared to x to measure relatedness. P* represents the frequency of the allele in the population (excluding all putative relatives of x). Using this equation, this statistical estimate of relatedness was calculated by comparing alleles shared between individuals to previously published population allele frequencies (van Oppen *et al*, 1997b) and to those derived for this study. It was necessary to include this information when calculating R to avoid the bias introduced by using the allele frequencies present in the broods sampled.

Table 2.3. Locus composition and mean expected heterozygosity for the five groups of loci used in this study. Group E contains the 6 loci that were used in Knight *et al* (1999)

Locus Group	Loci Included	Average H_E
A – 6 most polymorphic	Pzeb1-2, UNH002, UNH130, M11, DXTUCA-3	0.94
B – 6 least polymorphic	Pzeb3-5, UNH106, DXTUCA-3, UME003	0.75
C – all 11	All	0.84
D – 9 most polymorphic	Pzeb1-3, UNH002, UNH130, M11, DXTUCA-3, UNH106	0.90
E – 6 'original'	Pzeb1-5, UNH002	0.78

2.2.3.3 Rarefaction Analysis

In order to determine whether all loci contributed useful information during relatedness calculations a rarefaction analysis was also carried out. The method followed that of Smith *et al* (1997) and Piertney *et al* (1999). R values were calculated for a set of individuals using a single locus. Another locus was selected at random and R values were recalculated. This process continued until all loci had been used. The rarefaction

was repeated ten times with different random orders of loci. The mean absolute difference between the R values at each stage was used to determine the amount of information being contributed by each additional locus.

2.2.3.4 Simulations

Kinship 1.2 was also used to randomly generate 1000 pairs of full siblings, half siblings and unrelated individuals based on population allele frequency data at all 11 loci. This was repeated for the 6 loci used in Knight *et al* (1998) and the 9 most polymorphic loci for comparison.

2.3 Results

2.3.1 Linkage

The results from Genepop suggested linkage between two loci (Pzeb2 and DXTUCA-3, $p<0.05$ see Appendix 2.1). However, this result was rendered insignificant after Bonferoni correction for multiple tests; therefore all loci included in this study are regarded as unlinked.

2.3.2 Calculations of Pairwise R

The average heterozygosity of the locus group clearly has a strong effect on the calculated R values (Figure 2.1) – groups A, B and E all contain 6 loci, yet have produced very different results. The groups of loci with high average heterozygosities perform better as the range of R values generated for each category of relatedness was lower. Differences between the groups of loci vary depending on the category of relatedness examined; the groups behaved in a similar manner in the full-sib category, but differences were extremely marked when R was calculated for unrelated pairs. Mean R values were higher than the expected theoretical value of 0.25 in the half-sib category. The three locus groups with the highest heterozygosity exhibited the least amount of overlap in R values between relatedness categories (Table 2.3). Figure 2.2 shows the deviation from the expected theoretical R value exhibited for each relatedness category. Again, the locus groups with lower average heterozygosities performed relatively poorly. The groups which showed the smallest deviation from

expected were the two groups with the highest mean heterozygosities (groups C and D).

2.3.3 Rarefaction Analysis

Plotting the mean change in R with successive addition of loci (Figure 2.3) showed that the first few loci added have a strong influence on the R values calculated, but the effect of subsequent loci becomes progressively less until there appears to be little additional change in R after 9 loci.

2.3.4 Simulations

The distributions of the simulated R values can be seen in figure 2.4a, b and c. Means for all relatedness categories under all conditions used concurred with theoretical expected values (not shown). Overlap of R values between the relatedness categories was markedly reduced by increasing the number of loci used, although there was little difference between the distributions generated by using 9 and 11 loci.

Table 2.4. Total and component overlap between the four relatedness categories for each locus group. The three groups with the highest average heterozygosity exhibit the least amount of overlap in R values between relatedness categories.

Locus Group	Full sib/half sib overlap	Full sib/unrelated	Half sib/parent-offspring	Half sib/unrelated	Parent offspring /unrelated	Total Overlap
A	0.3447	0.0525	0.0436	0.0235	0	**0.4643**
B	0.4533	0	0.2663	0.3111	0	**1.0307**
C	0.2935	0	0.0886	0.0985	0	**0.4806**
D	0.3569	0	0.1224	0.0187	0	**0.498**
E	0.2789	0	0.1503	0.167	0	**0.5989**

Figure 2.1. The mean and range of R values generated for four relatedness categories ; a) Full sibs (61 pairwise comparisons), b) Parent/offspring (28 pairwise comparisons), c) Half sibs (33 pairwise comparisons), d) Unrelated pairs (18 pairwise comparisons). Number of loci and mean heterozygosity for each locus group are as follows A: n=6, H_E=0.94; B: n=6, H_E=0.75; C: n=11, H_E=0.84 ; D: n=9, H_E=0.90 ; E: n=6, H_E=0.78

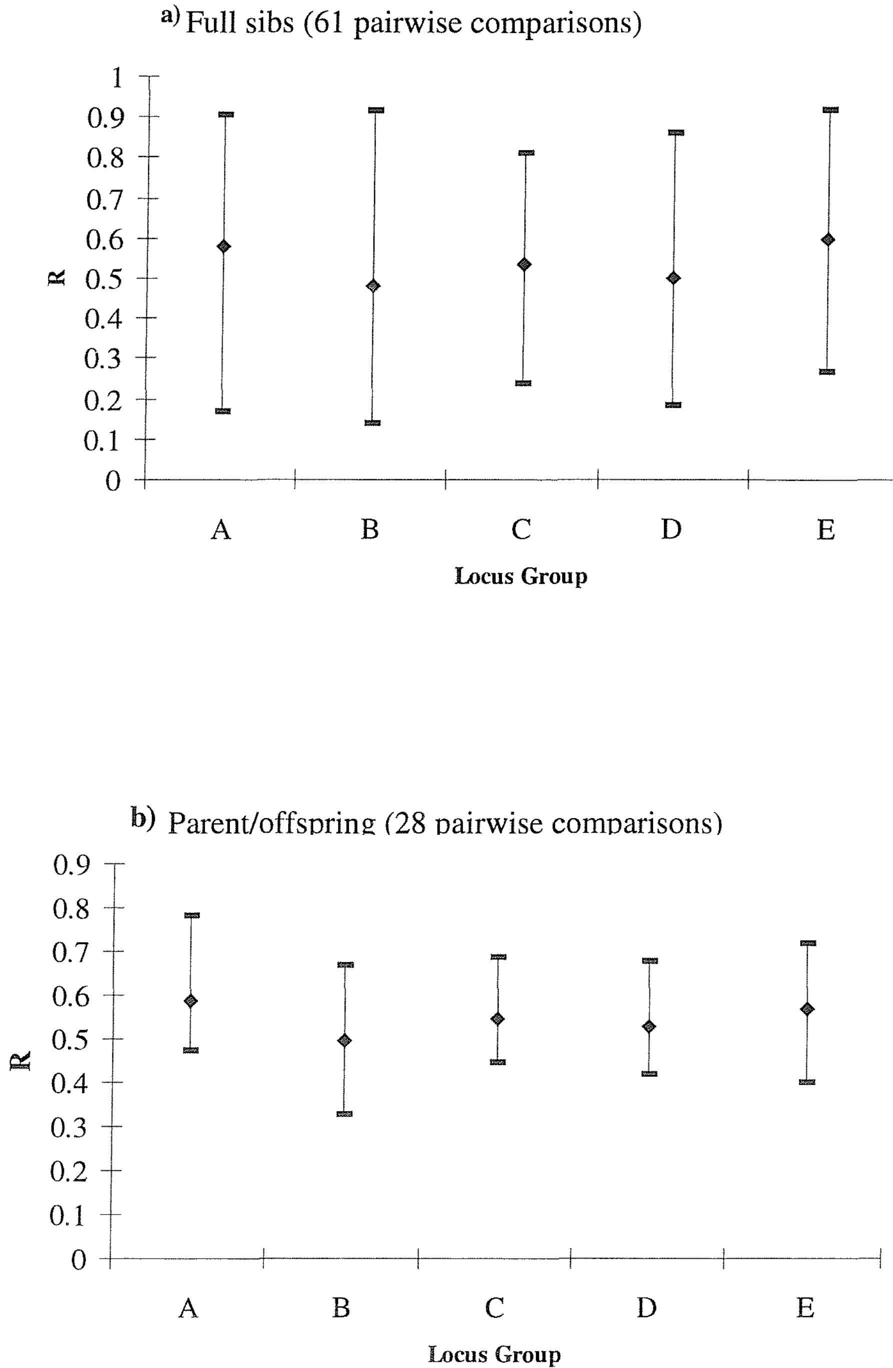


Figure 2.1 cont.

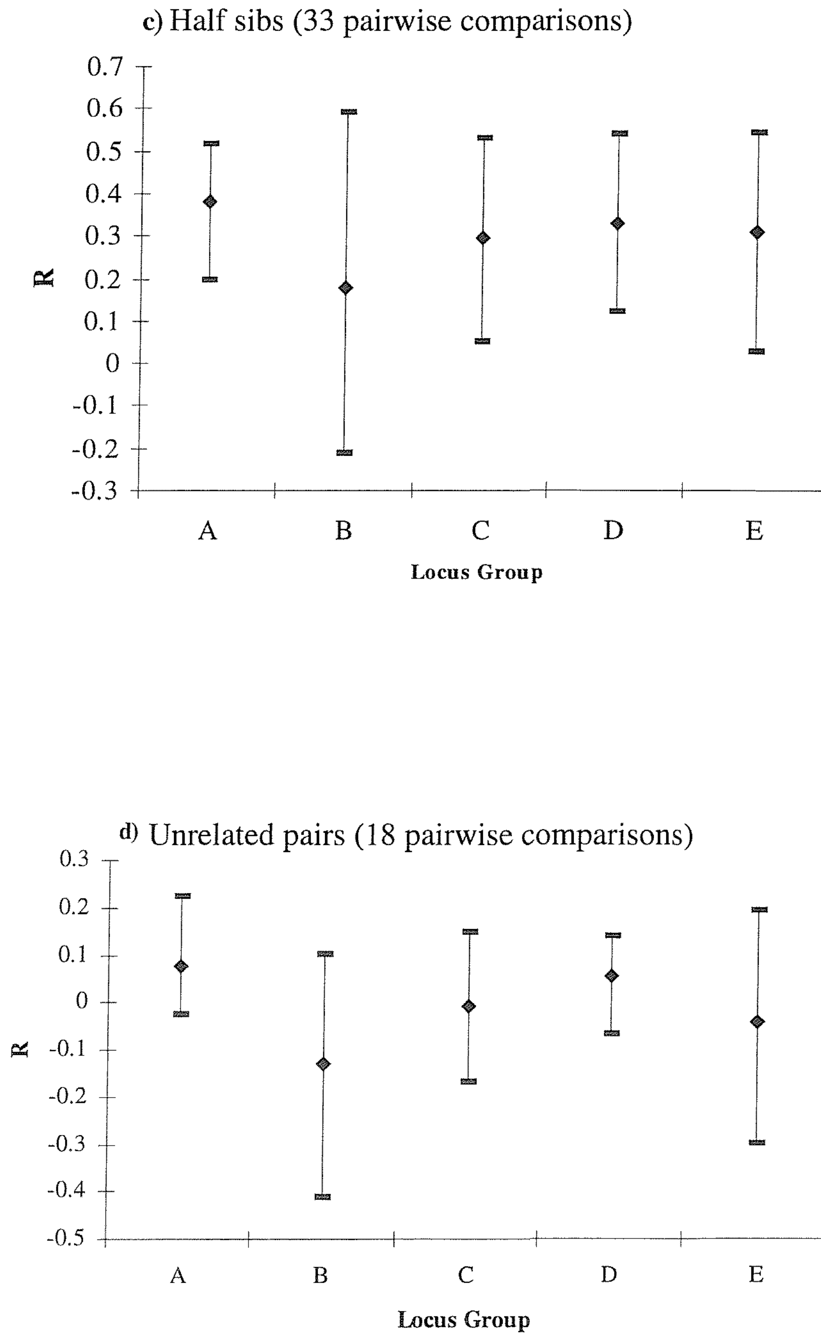


Figure 2.2. Deviation from expected R value calculated for known relatives for the four relatedness categories. Un=unrelated, fs=full sibs, hs=half sibs, po=parent/offspring. The two locus groups with consistently low deviation from expected values are the groups with the highest number of loci.

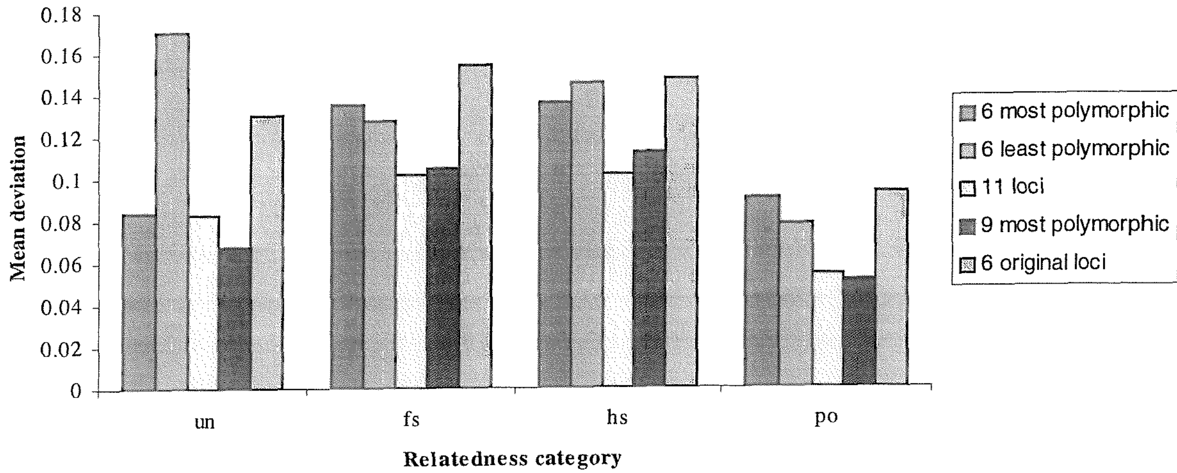


Figure 2.3. Rarefaction analysis for 11 loci shows that there is little improvement in relatedness estimation beyond 8 or 9 loci. Estimations were based on a single brood of *P. callainos* brood (full-sibs) and parents, and 10 random combinations of loci for each point. Means and standard deviations are shown. Further changes in R are minimal after 9 loci have been used.

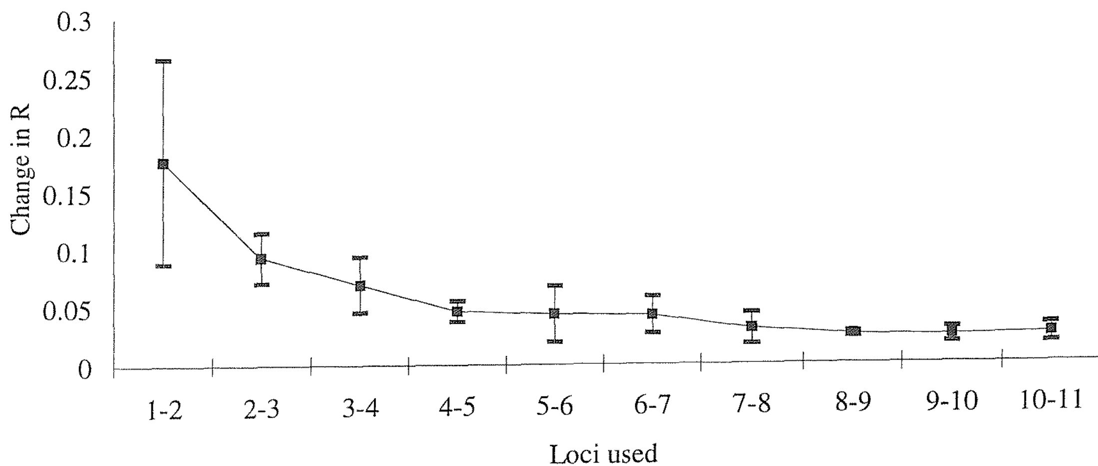
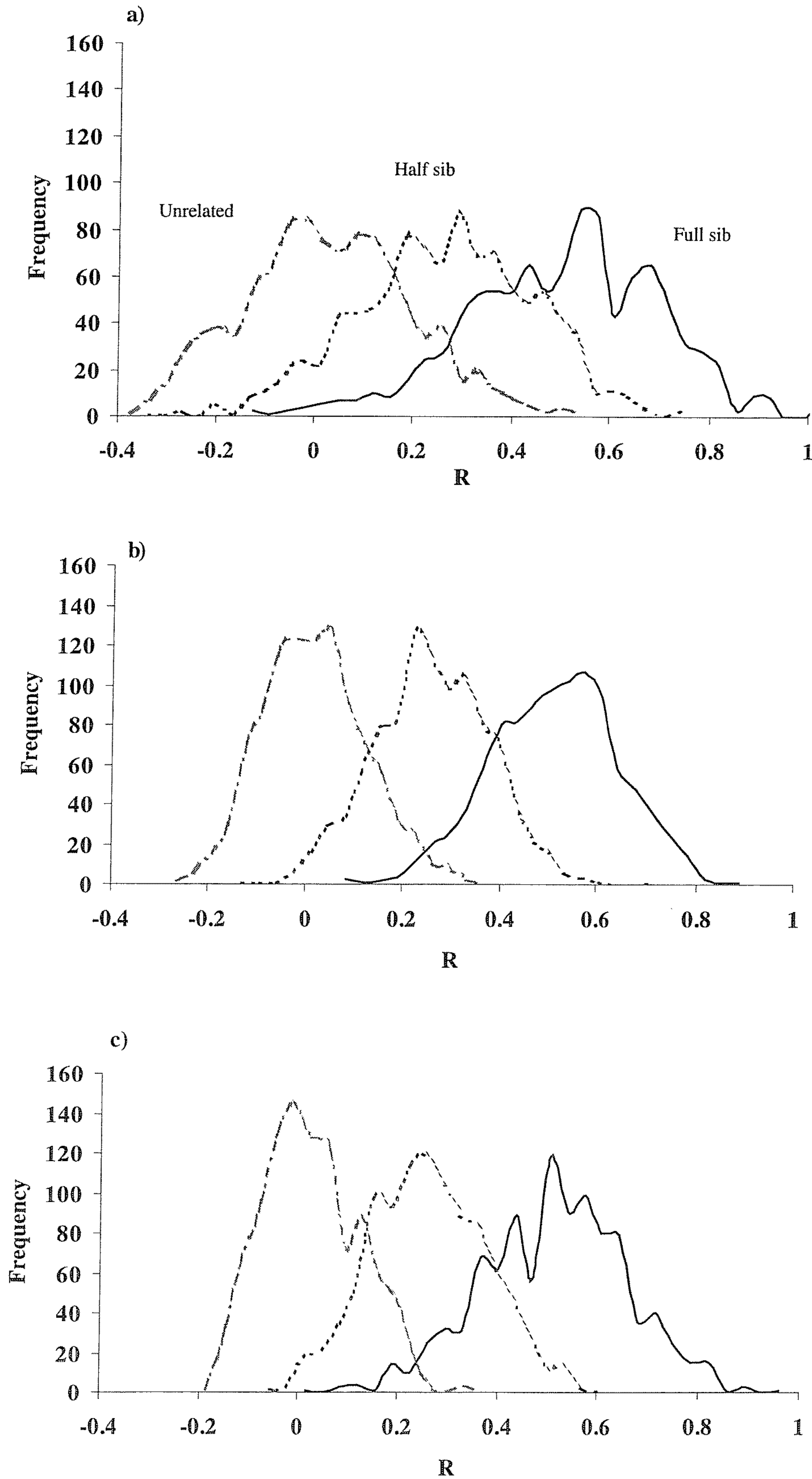


Figure 2.4. Distributions of 1000 R values for simulated full sibs, half sibs and unrelated pairs. Locus groups used in the simulations are as follows; a) Group E (6 loci, mean HE = 0.78), b) Group C (11 loci, mean HE = 0.84), c) Group D (9 loci, mean HE = 0.90).



2.4 Discussion

2.4.1 Optimisation of extra loci

The five extra loci (UME003, UNH130, UNH106, TmoM11 and DXTUCA-3) that had not previously been utilised in studies of relatedness in Malawi cichlids all amplified well, and despite conspicuous stutter could be scored unambiguously in all individuals. This is in spite of the fact that some had originally been isolated from cichlid species not native to Lake Malawi and indicates their usefulness in future studies of cichlids from this lake. A recently published study used UNH130 in a study of population divergence in pelagic cichlids (Shaw *et al*, 2000) has already demonstrated the applicability of this locus to another cichlid clade from Lake Malawi.

2.4.2 Relatedness Estimates

The results of this study indicate that including more loci, or using loci with higher heterozygosities provides better estimates of R. This is shown both by the tests carried out on known relatives and by the simulations. R values are both more accurate and precise when extra or more variable loci are used (in this case precision is measured by the range of R values generated for a particular category of relatedness). Obviously, a reduced range for each category gives greater discrimination between them. This can be seen clearly in the real and simulated datasets (Figures 2.1 and 2.4) and also in Table 2.4 which indicates the total amount of overlap for each locus group (known relative dataset). Interestingly the locus group that produces the least amount of overlap between relatedness categories is the one with the highest mean heterozygosity (group A, $H_E = 0.94$). However, two other locus groups that exhibit similar levels of overlap exhibit smaller deviations from the theoretically expected mean for each category (Figure 2.2). It appears that using smaller numbers of loci with very high levels of polymorphism may introduce bias into R values, possibly due to the influence of shared rare alleles on the statistic itself.

The performance of each group of loci varied between relatedness categories. The range of R values generated for full sibs pairs is relatively large compared to the

parent/offspring category. Both categories have an expected mean of R=0.5, but the range is much larger for full sibs as full sibs do not necessarily share any alleles at a given locus whereas parent/offspring pairs must have at least one allele per locus in common. In some cases, means for all groups deviate from expected values, for example the half-sib category. This is probably due to the two *M. zebra* females sharing alleles at four of the eleven loci, making their offspring apparently more related. In other cases, only the means of the most polymorphic groups deviate from expected. This is probably a result of the weight given to the sharing of rare alleles when calculating R – individuals are more likely to possess rare alleles if a locus is very variable. In all cases, the number of pairwise comparisons for each locus group/relatedness category combination is quite small. However, the general pattern of the results is in agreement with the results from the simulations.

The rarefaction analysis shows the same decrease in mean change in R as observed in other published studies (Smith *et al*, 1997, Piertney *et al*, 1999). The curve is not smooth as the rarefaction was not carried out exhaustively (i.e. all possible orders of loci). However, it does indicate that after data from 9 loci have been included, there is very little further change in R, this is again comparable with other published studies (Smith *et al*, 1997, Piertney *et al*, 1999) that reported no significant change in R after 8 and 10 loci respectively. The results suggest that little useful information is being contributed by the tenth and eleventh loci. With regard to future studies on natural populations, this means that pairs of individuals could be assigned to relatedness categories with approximately the same degree of accuracy when 9 loci rather than 11 have been used for statistical calculations.

In terms of numbers of loci used, this study is comparable with other published works (see Table 2.1). The majority of studies have used fewer than 11 loci, very few have used more than 15. The cost of typing hundreds of samples at just one extra locus can be considerable, therefore a prior indication of how many loci are required at the start of the study would be extremely useful. Simulations based on previously collected allele frequency data are particularly applicable to determining the minimum number of loci required as they can be carried out prior to typing the samples of interest.

The rarefaction analysis indicates that R values may not be greatly affected if only 9 loci are used rather than 11. The results from the tests on known relatives and the simulation data suggest that while there may be a small difference in the performance of groups with 9 and 11 loci, both behave in a similar fashion, and both outperform groups with fewer or less variable loci. In general, when considering which loci not to include, the results of this study indicate that it is preferable to include the most variable loci available, discounting those with the lowest expected heterozygosities. However, as demonstrated by the results from locus group A which contained just six of the most variable loci available, small numbers of highly polymorphic loci may produce biased estimates of relatedness. All groups of loci used in this study could discriminate between parent/offspring pairs and unrelated individuals (Table 2.4). The results suggest when these are the only relatedness categories under investigation (perhaps in studies of brood parasitism), it is possible to use smaller numbers of less polymorphic loci. With the exception of group A, most groups of loci could also distinguish between full sib and unrelated pairs, again suggesting locus number and heterozygosity is less important in this situation. However none of the groups used could distinguish half sibs from any other category of relatedness, although the more polymorphic groups resulted in less overlap between categories. It would probably be unfeasible to type samples at enough loci to discriminate between half sibs and other relatedness categories such as cousins, as also noted by Brookfield and Parkin (1993).

Queller and Goodnight's R is not the only relatedness estimator to have been developed for microsatellite data. Other estimators have been developed which employ different statistical methods, for example Lynch and Ritland (1999) and Ritland (1996), but have been less commonly used in published studies thus far. A search of the Web of Science database (http://wos.mimas.ac.uk) revealed that the paper detailing Queller and Goodnight's (1989) estimator has been cited 193 times since 1996 while that of Ritland (1996) has been cited 16 times. The paper describing the statistic of Lynch and Ritland (1999) has been cited 13 times since publication. This current study has dealt with only the frequently used R statistic; two published reviews have compared all relatedness estimators (van de Casteele *et al*, 2001; Lynch and Ritland, 1999). In both investigations it was found that the number of alleles per locus, the shape of the allele frequency distribution and the relatedness category (full-sib, half-sib, unrelated) influenced which estimator was more appropriate for a particular data set. This concurs

with the results obtained by this study, as both locus heterozygosity and relatedness category have an effect on the mean and range of R values generated for pairs of known pedigree.

2.4.3 Conclusions

In conclusion, this study has found that there appears to be an upper limit to the number of loci that can be considered 'useful' when calculating R, for the data set investigated this limit is around 9 loci. Factors such as locus variability and the category of relationship under investigation should influence the choice of loci used for calculating relatedness estimators such as Queller and Goodnight's R. Investigations into the number of loci such as those carried out in this study are advisable prior to beginning large-scale relatedness studies so that typing at excess loci is not carried out. This study would provide extremely useful background information for future field based studies of Malawi cichlids. Good estimates of relatedness would allow field based studies of the heritability of quantitative traits (Ritland, 2000) particularly male traits that have a possible influence on male reproductive success in groups such as the mbuna.

2.5 REFERENCES

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

Brookfield JFY and Parkin DT (1993) Use of single-locus DNA probes in the establishment of relatedness in wild populations. *Heredity* 70, 660-663

Chesser RK (1991) Gene diversity and female philopatry. *Genetics* 127 : 437-447

Favre L, Balloux F, Goudet J and Perrin N (1997) Female biased dispersal in the monogamous mammal *Crocidura russula*: evidence from field data and microsatellite patterns. *Proceedings of the Royal Society of London Series B* 264, 127-132

Goodnight KF and Queller DC (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology* 8, 1231-1234

Hamilton WD (1964) The genetical evolution of social behaviour I. *Journal of Theoretical Biology* 7: 1-16

Kays RW, Gittleman JL and Wayne RK (2000) Microsatellite analysis of kinkajou social organization. *Molecular Ecology* 9, 743-751

Kellogg KA, Markert JA, Stauffer JR and Kocher TD (1995) Microsatellite variation demonstrates multiple paternity in lekking fishes from Lake Malawi, Africa. *Proceedings of the Royal Society of London B* 260, 79-84

Knight ME, van Oppen MJH, Smith HL, Rico C, Hewitt GM and Turner GF (1999) Evidence for male biased dispersal in Lake Malawi cichlids from microsatellites. *Molecular Ecology* 8, 1521-1527

Lee W-J and Kocher TD (1996) Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *Journal of Fish Biology* 49, 169-171

Lynch M and Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753-1766

Parker A and Kornfield I (1996) Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi. *Environmental Biology of Fishes* 47, 345-352

Piertney SB, Maccoll ADC, Lambin X, Moss R and Dallas JF (1999) Spatial distribution of genetic relatedness in a moorland population of red grouse (*Lagopus lagopus scotius*). *Biological Journal of the Linnean Society* 68, 317-331

Pouyard L, Desmarais E, Chenuil A, Agnese JF and Bonhomme F (1999) Kin cohesiveness and possible inbreeding in the mouthbrooding tilapia *Sarotherodon melanotheron* (Pisces Cichlidae). *Molecular Ecology* 8, 803-812

Queller DC and Goodnight KF (1989) Estimating relatedness using molecular markers. *Evolution* 43, 258-275

Queller DC, Strassmann JE and Hughes CR (1993) Microsatellites and kinship. *Trends in Ecology and Evolution* 8, 285-288

Raymond M. & Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86, 248-249

Rico C *et al* (1992) Male reproductive tactics in the threespine stickleback – an evaluation by DNA fingerprinting. *Molecular Ecology* 1, 79-87

Ritland K (1996) Estimators for pairwise relatedness and inbreeding coefficients. *Genetical Research* 67, 175-186

Ritland K (2000) Marker inferred relatedness as a tool for detecting heritability in nature. *Molecular Ecology* 9, 1195-1204

Seutin *et al* (1991) Preservation of avian blood and tissue samples for DNA analysis. *Canadian Journal of Zoology* 69, 82-90

Shaw PW, Turner GF, Idid MR, Robinson RL and Carvaho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society of London B* 267: 2273-2280

Smith D, Meier T, Geffen E, Mech LD, Burch JW, Adams LG and Wayne RK (1997) Is incest common in gray wolf packs? *Behavioural Ecology* 8, 384-391

Surridge AK, Ibrahim KM, Bell DJ, Webb NJ, Rico C, and Hewitt GM (1999) Fine-scale genetic structuring in a natural population of European wild rabbits (*Oryctolagus cuniculus*). *Molecular Ecology* 8, 299-307

Taylor AC, Horsup A, Johnson CN, Sunnucks P and Sherwin B (1997) Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy nosed wombat *Lasorhinus krefftii*. *Molecular Ecology* 6 : 9-19

van de Casteele T, Galbusera P and Matthysen E (2001) A comparison of microsatellite based pairwise relatedness estimators. *Molecular Ecology* 10, 1539-1549

van Oppen MJH, Rico C, Deutsch JC, Turner GF and Hewitt GM (1997a) Isolation and characterisation of microsatellite loci in the cichlid fish *Pseudotropheus zebra*. *Molecular Ecology* 6, 387-388

van Oppen MJH, Turner GF, Rico C, Deutsch JC, Ibrahim KM, Robinson RL and Hewitt GM (1997b) Unusually fine scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London B* 264, 1803-1812

Worthington Wilmer J, Overall AJ, Pomeroy PP, Twiss SD and Amos W (2000) Patterns of paternal relatedness in British seal colonies. *Molecular Ecology* 9, 283-292

Zardoya R, Vollmer DM, Craddock C, Streelman GT, Karl S and Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their uses in resolving the phylogeny of cichlid fishes. *Proceedings of the Royal Society of London B* 263, 1589-1598

2.6 Chapter 2 Appendix

Table 2.5. P values returned by Genepop for tests of genotypic linkage disequilibrium between loci (null hypothesis of no significant linkage between loci).

Locus 1	Locus 2	P-value	S.E.
Pzeb4	Pzeb5	0.87319	0.00168
Pzeb4	UNH002	0.42692	0.01200
Pzeb5	UNH002	0.47226	0.00373
Pzeb4	Pzeb2	1.00000	0.00000
Pzeb5	Pzeb2	0.75682	0.00294
Pzeb4	Pzeb1	1.00000	0.00000
Pzeb5	Pzeb1	1.00000	0.00000
UNH002	Pzeb1	0.51732	0.00372
Pzeb2	Pzeb1	1.00000	0.00000
Pzeb4	Pzeb2	1.00000	0.00000
Pzeb5	Pzeb2	0.86436	0.00492
UNH002	Pzeb2	0.25712	0.00315
Pzeb2	Pzeb2	0.05251	0.00468
Pzeb1	Pzeb2	0.69566	0.00976
Pzeb4	UNH130	0.76752	0.00943
Pzeb5	UNH130	1.00000	0.00000
UNH002	UNH130	0.30237	0.00302
Pzeb2	UNH130	1.00000	0.00000
Pzeb1	UNH130	1.00000	0.00000
Pzeb2	UNH130	1.00000	0.00000
Pzeb4	UME003	0.79180	0.00835
Pzeb5	UME003	1.00000	0.00000
UNH002	UME003	0.24341	0.00333
Pzeb2	UME003	1.00000	0.00000
Pzeb1	UME003	1.00000	0.00000
Pzeb2	UME003	1.00000	0.00000
UNH130	UME003	1.00000	0.00000
Pzeb4	DXTUCA-3	1.00000	0.00000
Pzeb5	DXTUCA-3	0.39889	0.00289
UNH002	DXTUCA-3	1.00000	0.00000
Pzeb2	DXTUCA-3	1.00000	0.00000
Pzeb1	DXTUCA-3	1.00000	0.00000
Pzeb2	DXTUCA-3	0.03131	0.00250
UNH130	DXTUCA-3	1.00000	0.00000
UME003	DXTUCA-3	1.00000	0.00000
Pzeb4	UNH106	1.00000	0.00000
Pzeb5	UNH106	0.24656	0.00000
UNH002	UNH106	1.00000	0.00270
Pzeb2	UNH106	1.00000	0.00000
Pzeb1	UNH106	1.00000	0.00000
Pzeb2	UNH106	0.78103	0.00000
UNH130	UNH106	1.00000	0.00832
UME003	UNH106	1.00000	0.00000
DXTUCA-3	UNH106	1.00000	0.00000
Pzeb4	TmoM11	0.19050	0.00955
Pzeb5	TmoM11	1.00000	0.00000

Locus 1	Locus 2	P-value	S.E.
UNH002	TmoM11	1.00000	0.00000
Pzeb2	TmoM11	1.00000	0.00000
Pzeb1	TmoM11	1.00000	0.00000
Pzeb2	TmoM11	1.00000	0.00000
UNH130	TmoM11	1.00000	0.00000
UME003	TmoM11	1.00000	0.00000
DXTUCA-3	TmoM11	1.00000	0.00000
UNH106	TmoM11	1.00000	0.00000

Table 2.6. Genotypes of the *M. zebra* females (1 and 4) and the *M. zebra* male (10) and offspring at the 11 lcoi. Data from Pzeb1-5 and UNH002 taken from Knight *et al* (2000).

ID	PZEB4	PZEB5	UNH2	PZEB1	PZEB2	PZEB3	M11	UNH130	UME003	DXT3	UNH106
1	127/131	129/129	207/223	170/184	265/265	314/320	207/221	212/224	209/243	131/133	149/153
10	123/123	129/137	207/215	164/170	263/263	320/322	213/216	208/228	203/225	117/133	159/169
1B	123/127	129/129	207/215	170/184	263/265	314/320	207/216	212/228	209/225	117/133	153/169
1D	123/127	129/137	207/215	170/184	263/265	314/320	207/216	224/228	209/225	131/133	149/159
1F	123/131	129/129	207/215	170/184	263/265	314/320	207/213	212/228	209/225	117/133	149/169
4	123/131	137/137	211/293	194/228	245/263	314/316	207/216	174/220	203/235	127/131	147/165
4A	123/131	129/137	207/211	164/194	263/263	314/322	207/216	220/228	203/235	127/133	147/159
4B	123/123	137/137	207/293	170/228	245/263	316/320	213/216	174/208	225/235	117/127	165/169
4D	123/123	129/137	207/211	170/228	263/263	314/320	213/216	220/228	203/225	117/131	147/169
4G	123/123	129/137	215/293	170/228	245/263	316/322	213/216	208/220	203/225	117/131	159/165
4H	123/131	129/137	211/215	170/194	263/263	314/322	207/216	174/208	203/225	131/133	147/169
4I	123/123	137/137	207/293	170/194	245/263	316/322	207/216	174/208	203/235	131/133	147/159
4J	123/131	129/137	207/211	170/194	263/263	314/320	216/216	174/228	225/235	117/131	159/165
4L	123/123	137/137	215/293	164/228	263/263	314/322	207/213	208/220	203/203	117/131	147/169
4N	123/123	129/137	211/215	170/228	263/263	314/320	207/216	208/220	225/235	131/133	147/159
4R	123/131	129/137	207/293	170/194	245/263	314/320	213/216	174/228	203/225	127/133	147/169
4T	123/123	137/137	207/211	170/228	245/263	316/320	207/213	208/220	225/235	127/133	165/169

Table 2.7. Genoypes of the *M. callainos* individuals (5- mother, 13 – sire) at the 11 loci used in this study. Data from Pzeb1-5 and UNH002 taken from Knight *et al* (2000).

ID	PZEB4	PZEB5	UNH2	PZEB2	PZEB1	PZEB3	UNH130	UME003	DXT3	UNH106	M11
5	123/133	123/129	211/223	251/279	220/252	316/330	212/226	225/227	141/141	161/163	198/211
13	133/133	123/129	197/211	245/265	174/210	318/328	226/226	205/227	131/139	145/161	208/208
5B	133/133	129/129	197/211	245/279	174/220	318/330	226/226	205/227	139/141	161/163	208/211
5L	133/133	129/129	211/211	265/279	174/252	318/330	212/226	205/225	131/141	145/163	198/208
5M	133/133	123/129	211/211	251/265	210/220	318/330	226/226	225/227	139/141	161/161	198/208

Table 2.8. Allele frequencies at the 11 loci in *M. zebra*. Data for Pzeb1-5 and UNH002 taken from van Oppen *et al* (1997b).

PZEB4		PZEB5		UNH002		PZEB1		PZEB2		PZEB3		M11		UNH130		UME003		DXT3		UNH106	
123	0.297	121	0.020	174	0.102	134	0.005	204	0.005	314	0.204	199	0.028	174	0.086	193	0.036	103	0.010	127	0.033
127	0.317	123	0.029	179	0.015	136	0.005	206	0.005	316	0.032	201	0.014	176	0.010	197	0.012	109	0.010	129	0.078
129	0.010	127	0.015	183	0.005	138	0.005	226	0.005	318	0.059	202	0.056	178	0.010	201	0.012	111	0.021	131	0.022
131	0.094	129	0.794	189	0.005	142	0.005	227	0.049	320	0.425	207	0.042	182	0.048	203	0.048	113	0.021	133	0.022
133	0.168	131	0.010	193	0.010	144	0.020	229	0.029	322	0.177	209	0.014	184	0.029	205	0.012	117	0.082	145	0.233
137	0.064	137	0.132	195	0.010	148	0.010	235	0.029	324	0.016	211	0.069	186	0.010	209	0.048	121	0.052	147	0.033
139	0.020			197	0.041	150	0.055	239	0.029	326	0.038	212	0.056	190	0.010	211	0.048	123	0.052	149	0.078
141	0.005			199	0.015	152	0.010	241	0.019	328	0.022	213	0.083	198	0.010	213	0.071	125	0.021	151	0.022
143	0.010			201	0.015	156	0.010	243	0.015	330	0.011	214	0.014	202	0.076	215	0.048	127	0.144	153	0.089
145	0.015			203	0.020	158	0.015	245	0.019	334	0.011	215	0.097	204	0.095	217	0.024	129	0.021	155	0.033
				205	0.010	160	0.025	247	0.015	338	0.005	216	0.042	206	0.086	219	0.024	131	0.196	157	0.144
				207	0.117	162	0.035	249	0.019			217	0.056	208	0.019	221	0.048	133	0.206	159	0.056
				209	0.036	164	0.100	251	0.058			218	0.042	210	0.029	223	0.060	135	0.062	161	0.011
				211	0.046	166	0.060	253	0.005			219	0.097	212	0.029	225	0.131	137	0.093	163	0.011
				213	0.005	168	0.030	255	0.019			220	0.028	214	0.029	227	0.024	161	0.010	165	0.067
				215	0.026	170	0.070	257	0.058			221	0.042	216	0.010	229	0.095			169	0.056
				217	0.020	172	0.030	259	0.039			222	0.042	218	0.076	231	0.083			173	0.011
				219	0.020	174	0.020	261	0.044			223	0.028	220	0.029	233	0.060				
				221	0.020	176	0.060	263	0.136			225	0.028	222	0.048	235	0.024				
				223	0.015	178	0.040	265	0.087			226	0.028	224	0.095	237	0.036				
				225	0.026	180	0.015	267	0.073			227	0.042	226	0.067	241	0.036				
				227	0.056	182	0.090	269	0.029			229	0.028	228	0.019	243	0.012				
				229	0.031	184	0.020	271	0.078			231	0.014	230	0.010	245	0.012				
				231	0.041	186	0.020	273	0.029			233	0.014	232	0.019						
				233	0.056	188	0.015	275	0.039					234	0.038						
				235	0.077	190	0.005	277	0.034					240	0.010						

PZEB4		PZEB5		UNH002		PZEB1		PZEB2		PZEB3		M11		UNH130		UME003		DXT3		UNH106	
				237	0.020	192	0.050	279	0.015					244	0.010						
				239	0.031	194	0.010	281	0.015												
				241	0.026	196	0.010	286	0.005												
				243	0.026	200	0.015														
				245	0.010	202	0.005														
				247	0.020	204	0.005														
				249	0.010	206	0.005														
				251	0.010	210	0.025														
				293	0.005	212	0.015														
						214	0.010														
						216	0.005														
						218	0.005														
						220	0.005														
						222	0.005														
						224	0.005														
						226	0.010														
						228	0.005														
						230	0.005														
						234	0.010														
						238	0.005														
						240	0.005														
						242	0.010														

Table 2.9. Allele frequencies of the 11 loci in *M. callainos*. Data for Pzeb1-5 and UNH002 taken from van Oppen *et al* (1997b).

PZEB4		PZEB5		UNH002		PZEB2		PZEB1		PZEB3		UNH130		UME003		DXT3		UNH106		M11	
123	0.100	123	0.559	174	0.013	237	0.016	138	0.004	314	0.305	176	0.021	199	0.011	107	0.011	141	0.022	198.000	0.014
131	0.439	129	0.441	189	0.017	239	0.029	150	0.030	316	0.114	178	0.011	203	0.011	115	0.065	145	0.011	200.000	0.014
133	0.457			191	0.042	241	0.012	158	0.009	318	0.150	180	0.011	205	0.011	119	0.022	149	0.011	202.000	0.014
137	0.004			193	0.021	243	0.012	160	0.026	320	0.314	182	0.053	207	0.011	123	0.022	151	0.033	203.000	0.014
				195	0.029	245	0.012	162	0.004	322	0.045	184	0.021	213	0.022	125	0.011	153	0.055	204.000	0.057
				197	0.042	247	0.041	164	0.026	324	0.009	190	0.011	215	0.011	127	0.109	155	0.132	205.000	0.014
				199	0.033	249	0.029	166	0.004	326	0.005	192	0.011	217	0.011	129	0.065	157	0.022	207.000	0.029
				201	0.013	251	0.037	168	0.038	328	0.018	194	0.043	219	0.067	131	0.130	159	0.330	208.000	0.143
				203	0.058	253	0.041	170	0.047	330	0.018	196	0.011	221	0.011	133	0.054	161	0.187	209.000	0.014
				205	0.054	255	0.053	172	0.026	332	0.018	198	0.043	223	0.056	135	0.065	163	0.055	210.000	0.014
				207	0.067	257	0.020	174	0.038	342	0.005	200	0.011	225	0.167	137	0.087	165	0.033	211.000	0.014
				209	0.054	259	0.098	176	0.004			202	0.011	227	0.100	139	0.065	167	0.011	212.000	0.071
				211	0.042	261	0.049	178	0.017			204	0.032	229	0.100	141	0.152	169	0.088	213.000	0.014
				213	0.021	263	0.078	180	0.021			206	0.053	231	0.044	143	0.022	177	0.011	214.000	0.014
				215	0.025	265	0.127	182	0.030			208	0.011	233	0.033	145	0.022			215.000	0.057
				217	0.058	267	0.107	184	0.021			210	0.043	235	0.067	147	0.033			216.000	0.029
				219	0.050	269	0.102	186	0.013			212	0.032	237	0.044	149	0.054			218.000	0.071
				221	0.088	271	0.029	188	0.026			214	0.011	239	0.044	151	0.011			220.000	0.014
				223	0.046	273	0.041	190	0.038			216	0.021	241	0.089					223.000	0.014
				225	0.054	275	0.029	192	0.030			220	0.011	243	0.067					225.000	0.043
				227	0.033	277	0.008	194	0.043			222	0.064	245	0.022					227.000	0.114
				229	0.063	279	0.020	196	0.056			224	0.064							229.000	0.086
				231	0.033	281	0.008	198	0.030			226	0.032							231.000	0.086
				233	0.013			200	0.034			228	0.117							232.000	0.014
				235	0.004			202	0.017			230	0.074							233.000	0.014

PZEB4		PZEB5		UNH002		PZEB2		PZEB1		PZEB3		UNH130		UME003		DXT3		UNH106		M11	
				237	0.013			204	0.026			234	0.021							248.000	0.014
				239	0.004			206	0.030			236	0.064								
				241	0.008			208	0.026			238	0.032								
				243	0.000			210	0.013			240	0.032								
				245	0.004			212	0.034			242	0.011								
								214	0.004			244	0.011								
								216	0.013			246	0.011								
								218	0.030												
								220	0.038												
								222	0.034												
								224	0.013												
								226	0.009												
								228	0.013												
								230	0.004												
								232	0.013												
								234	0.009												
								236	0.013												
								240	0.013												
								242	0.013												
								244	0.009												
								246	0.009												
								252	0.004												

3 Allopatric Populations and Tests of Mate Choice

3.1 Chapter Note

This chapter describes research carried out by M.E. Knight, C. J. Allender and R. Poole. M.E. Knight was responsible for experimental design and fish maintenance. C. J. Allender carried out the majority of the genotyping. R. Poole carried out some genotyping as part of an undergraduate project.

3.2 Introduction

The existence of allopatric populations has presented problems in the definition of species. Mayr's (1942) Biological Species Concept uses reproductive isolation between populations to describe species but by definition it applies to sympatric populations and not to populations that are geographically isolated from one another. The Recognition Concept of Paterson (for example 1980) described species as groups of individuals sharing a Specific Mate Recognition System or SRMS. The SRMS is a collective term for the mechanisms whereby individuals perceive potential mates. Reproductive isolation arises when two populations have diverged in the SRMS. In the case of allopatric populations, this is proposed to be due to different selective pressures operating in different environments. However, the Recognition Concept does not address issues of hybrid inviability or sterility (Coyne *et al*, 1988) and in practice it is difficult to assess divergence in the SMRS among allopatric populations.

The taxonomic status of allopatric populations is especially relevant to the Malawi species flock. Current species richness estimates are in the range of 500 (van Oppen *et al*, 1997) to over 1000 (Turner, 1999). The mbuna group of the Malawi flock is composed of relatively small rock dwelling cichlids. Mbuna are sexually dimorphic, males having bright breeding colours and exaggerated fins and females having relatively cryptic coloration. The males are usually highly territorial, with females visiting male territories in order to spawn. A great deal of research has focused on mbuna species as they occupy easily accessible shallow water habitats and populations are easily maintained in aquariums. Studies of mbuna populations in the wild and in the

laboratory have provided insight into evolutionary mechanisms that generate and maintain diversity (for example van Oppen *et al*, 1997 and 1998; Arnegard *et al*, 2000; Danley *et al*, 2001; Knight *et al*, 1998; Knight and Turner, 1999). In an initial survey, Ribbink *et al* (1983) recorded 196 populations differing in male colour pattern, which they regarded as separate species. However, only 51 of these colour forms occur sympatrically, meaning that they can be classified under the BSC as described above (Turner *et al*, 2001). Species richness estimates for Lake Malawi obviously depend on whether allopatric colour forms really do represent good species.

The question of reproductive isolation in mbuna species has already received much attention. Species that appear to be reproductively isolated in nature can be crossed to produce viable inter-specific and even inter-generic hybrids (e.g. Crapon de Caprona and Fritzsch, 1984; Albertson and Kocher, 2001; Knight *et al*, 1998), although this only occurs under laboratory conditions. Reproductive isolation between sympatrically occurring species has also been demonstrated using laboratory tests (Knight *et al*, 1998) and by comparing microsatellite allele frequencies in wild populations (van Oppen *et al*, 1998). Field based direct tests of reproductive isolation would require unethical large scale transplants of individuals.

The problem of the taxonomic status of allopatric populations is particularly pertinent to the mbuna, as many mbuna 'species' exist as geographic races that differ only in male colour pattern (Ribbink *et al*, 1983). Neighbouring allopatric populations have been shown to be genetically distinct from one another (van Oppen *et al*, 1997; Markert *et al*, 1999; Arnegard *et al*, 1999 Danley *et al* 2001). Ribbink *et al* (1983) provided the first survey of mbuna species, classifying them according to male colour pattern, and reported the existence of 196 species. However, Turner *et al* (2001) point out inconsistencies in the assignment of specific status to certain populations, and go on to suggest that there may be as few as 51 valid species when only sympatrically occurring male colour forms are taken into account. Stauffer *et al* (1997) redescribed several mbuna species using a slightly different approach involving morphology as well as male colour patterns. Specific status was assigned to allopatric populations that were previously regarded as conspecific. Turner (1999) maintained that promotion of allopatric populations to species should not take place without further investigations of breeding behaviour.

The current study examines the mate choice of two species of mbuna, *Metriaclima callainos* and *M. zebra*. Both of these species exist as geographic races in different areas of the lake. Male colour pattern varies between races as can be seen in Figure 3.1. The locations of the populations used in this experiment are shown in Figure 3.2. It is unlikely that there is any gene flow between them as Bouteillon (1998) found high genetic differentiation and extremely low levels of migration using microsatellite based F_{ST} statistics. A demonstration of reproductive isolation could affect their current conspecific status. To this end, individuals from each race were allowed to mix freely in experimental tanks. Eggs were removed from brooding females and the paternity of offspring was determined using microsatellites.

Figure 3.1. a) Male representatives of the Nkhata Bay race (Nkhata Bay population) and the Ruarwe population of *M. zebra.* The main differences are the yellow throat and black submarginal band in the dorsal fin of the Chilumba male and the dark cheek of the Nkhata Bay male. Photos taken by Ad Konings and reproduced with kind permission

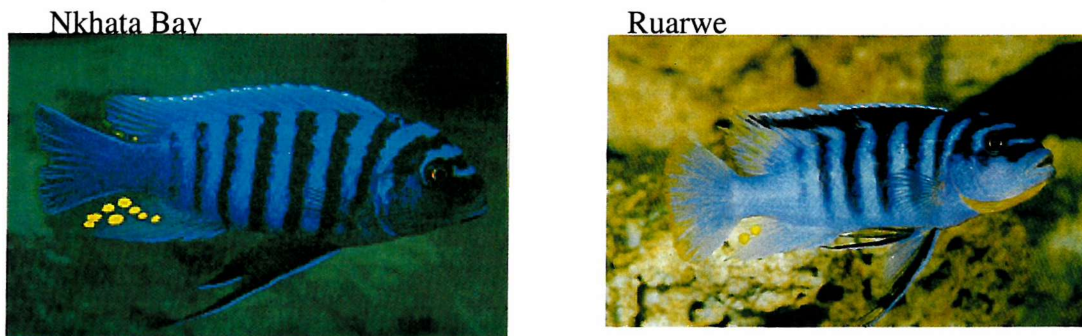
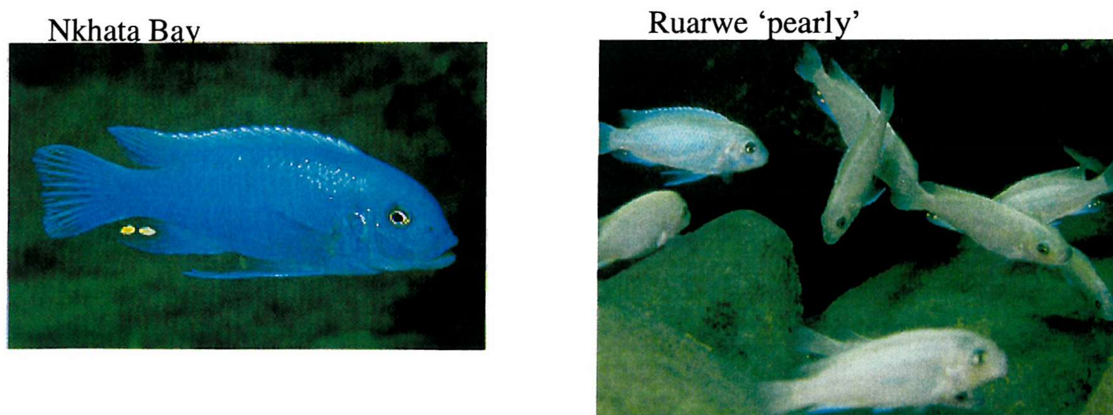


Figure 3.1 b. Male *M. callainos* of the Nkhata Bay and Ruarwe 'Pearly' races. Photos taken by Ad Konings and reproduced with kind permission



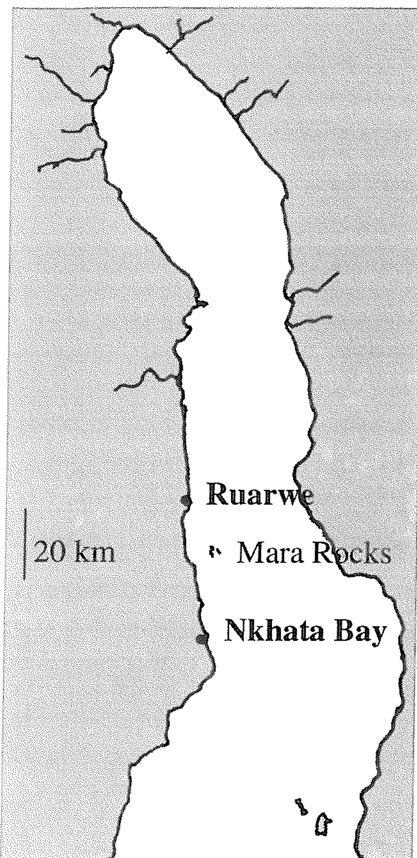


Figure 3.2. Map of the northern half of Lake Malawi showing the location of the populations used in this study (shown in bold).

3.3 Methods

3.3.1 Experimental Design

Laboratory tests of mate choice were performed on captive fish of the two species (*M. callainos* and *M. zebra*). For each species, one male and three to four females of each race were placed in experimental tanks measuring 61 x 46 x 153 cm. Clips from the soft rayed posterior of the dorsal fin were taken from males prior to the start of the experiment. The *M. zebra* females were distinguished by using different colour morphs for each race. Female coloration is polymorphic in many mbuna species and this occurs in both the races used in this experiment. The colour pattern of females of *M. callainos* differs in a similar way to the male coloration, which also allows females of the two races to be distinguished. Spawning sites were provided in each tank in the form of rocks and terracotta pots. The fish were then left to breed freely. Mouthbrooding females were isolated, stripped of their eggs and a fin clip was taken. Eggs were either preserved immediately or left to develop in a hatchery (in this case, fin clips of the fry were taken once they had reached 5 cm in length. With the exception of the first

repetitions, females were returned to the tanks. This was done to maintain female densities and reduce male aggression. All genetic samples were preserved in DMSO buffer (Seutin *et al*, 1991)

The *M. callainos* experiment consisted of two repetitions, the first running from December 1996 to April 1997 and the second from April 1997 to November 1998. All of the adults used were wild caught. The *M. zebra* experiment consisted of three repetitions, the first from December 1996 to June 1997, and the second and third from June 1997 to November 1998. Adults used in the first repetition were wild caught, but those used in the second and third were the F1 generation of laboratory stock broods. Maintenance of experimental tanks followed Knight *et al* (1998). Tanks were kept on a 12:12 light-dark cycle, and were illuminated with white fluorescent bulbs.

3.3.2 Genetic Analysis

DNA was extracted using a proteinase K/chloroform protocol as described in Chapter 2 (Rico *et al*, 1992). In most cases the entire egg was homogenised and extracted, although if the larva was visible, the yolk sac was removed and discarded. Samples were resuspended in 200 μL 10mM Tris HCl. Due to the low amounts of DNA likely to be present in young eggs, DNA from these was resuspended in 100 μL.

Loci used for paternity assignment were Pzeb1 (van Oppen *et al*, 1997) and UNH130, UME003, DXT3, TmoM11and UNH106 (see Table 2.2, Chapter 2 for characteristics and original references to these loci). It was decided that the status of hybrid broods should be confirmed by at least two loci. Rep 1 for each experiment was originally analysed using Pzeb1 (R. Poole, Southampton University, undergraduate project 1998). Some of the results for the two reps were unclear, and required typing at further loci.

PCR protocols followed those described in Chapter 2 of this report. PCR products were sized using an ABI 377 DNA sequencer with a TAMRA labelled GeneScan 350 internal size standard. Where samples failed to yield a PCR product or produced a great deal of stutter, a variety of means were used to attempt to correct the problem. Template DNA and primer concentration were both reduced and increased, magnesium levels in the PCR mix were varied and annealing temperatures and the number of PCR

cycles performed were also varied. Different brands of *Taq* polymerase were also used, as was a proprietary kit for problem PCRs (FailSafe kit, Microzone Ltd) which contained various buffers and PCR enhancers (BSA and betaine). In an attempt to rid the DNA extract of PCR inhibiting compounds, aliquots of certain extracts were cleaned using PCR purification columns (Qiagen). Products were also run on the sequencer undiluted (usually they are diluted 20-50 fold before being loaded).

Allele peaks were identified by eye from the electropherogram output of the sequencer. Paternity of the offspring was determined manually rather than through the use of a software package by comparing alleles present in each offspring to the maternal genotype and that of the potential fathers.

3.4 Results

3.4.1 Assortative Mating

Genotypes of parents and offspring can be found in the appendix at the end of this chapter. Not all broods have been amplified at all loci, sometimes because it was not necessary and sometimes because only certain loci would amplify. The results of the two experiments can be seen in Table 3.1 below. It is clear that hybridisation between the two *M. zebra* races has occurred – in Rep 2 of this experiment, 20 out of 25 individuals in one brood from a Ruarwe female (female C 995) were sired by the Nkhata Bay male. The paternity of the remaining five individuals could not be ascertained with the two loci used. Hybrid broods may be present in the other repetitions, as results were inconclusive in a few cases due to PCR problems. However, in the other reps, if hybrids are present then the broods are the result of multiple paternity, as the sire of at least some of the individuals was clear.

Hybridisation may have occurred in the *M. callainos* experiment as well, but again results are not conclusive due to PCR problems. In both of the potential hybrid *M. callainos* broods, hybrids were identified at locus Pzeb 1, but genotypes at other loci did not agree. Certainly the majority of the broods appear to be the product of assortative mating.

3.4.2 PCR Problems

In many cases in this experiment, amplification of PCR product from DNA extracted from eggs was not straightforward. Despite the measures described in the Methods section, PCR products were not obtained for many individuals. An extreme case of this can be seen in Rep 2 of the *M. zebra* experiment where no product was obtained for the majority of broods. In other cases the product was very 'messy' with large amounts of stutter or even more than two possible allelic peaks. Quite often, the yield of product was so low that allele sizes could not be easily determined using the sequencer. The quality and concentration of DNA extractions were tested on a 0.9% agarose gel stained with ethidium bromide. However, the yield of DNA extracted from eggs was often too low to be detected by this method. The presence of DNA in the extracts remained unknown in these cases.

Table 3.1. Summary of results for all repetitions of both experiments. Population of origin of each female is as follows. RP – Ruarwe, NB – Nkhata Bay (both *M. callainos*), C – Ruarwe, NB – Nkhata Bay (both *M. zebra*). Numbers in brackets in 'Loci' column indicate the number of individuals amplifying at that locus, where this varied greatly from locus to locus. Notes on possible hybrids between the two races are as follows :

1. Hybrids identified at Pzeb1, most of these refuted by other loci. Amplification messy and inconclusive.
2. 6 hybrids according to Pzeb1, 4 of these contradicted by UNH130. Others messy.
3. Hybrids identified at Pzeb1, but unable to amplify at any other locus.
4. Hybrids identified at UNH130, but UME003 produced inconclusive results and did not confirm.
5. 20 definite hybrids. Not clear from the 2 loci whether the remaining 5 are the result of multiple paternity.
6. Messy, female sample failed to amplify.

Female	Brood Size	N amplified	Loci used	Ambiguities	Hybrids
M. callainos Rep 1					
RP 399	3	3	Pzeb1, UNH130	No	No
RP 401	4	1	UNH130	No	No
RP 420	10	10	Pzeb1	No	No
NB 318B	5	5	Pzeb1	No	No
NB 421	13	11	Pzeb1, UNH130, DXT3, UME003, M11	Yes	*? See note 1*
NB 447	11	3	UNH130, DXT3, Pzeb1	No	no

Female	**Brood Size**	**N amplified**	**Loci used**	**Ambiguities**	**Hybrids**
M. callainos Rep 2					
RP 470	16	14	Pzeb1, UNH130	No	no
RP 564	14	12	Pzeb1, UNH130	No	no
NB 459	14	9	UNH130, Pzeb1, DXT3, M11	Yes	*? See note 2*
NB 505	3	3	Pzeb1, UNH130	No	No
NB 750	7	6	UNH130	No	No
NB 767	7	7	UNH130	No	No
NB 1292	11	0	UNH130, DXT3, M11	?	?
M. zebra Rep1					
C 313	15	14	Pzeb1	No	no
C 542	57	27	Pzeb1, UNH130 (29)	?	*? 1 possible – not confirmed*
C 318A	10	10	Pzeb1 UNH130 (5)	No	No
NB 406	10	5	Pzeb1	No	*Yes (5) but only 1 locus used*
NB 469	18	16	Pzeb1, DXT3	No	No
NB 536	21	13	Pzeb1	Yes	*? See note 3*
NB 553	4	4	UNH130	No	no
M. zebra Rep 2					
C 785	16	0	All tried	?	
NB 802	22	0	All tried	?	
NB 825	16	0	All tried	?	
NB 882	22	0	All tried	?	
C 905	9	8	UNH130 (3), UME003 (8)	Yes	*? See note 4*
NB 925	25	0	All Tried	?	
C 995	25	25	UNH130, DXT3	Yes	*Yes (20) see note 5*
NB 1021	14	0	UNH130, DXT3, UME003, M11	Yes	inconclusive
M. zebra **Rep 3**					
C 843	9	5	UNH130, UME003	Yes	*? See note 6*
C 854	26	17	UME003	No	No
C 915	31	29	UNH130, DXT3	No	No
NB 946	5	5	UNH130, DXT3	No	No
NB 978	16	9	UME003, M11	Yes	inconclusive
NB 1036	12	8	UME003	Yes	inconclusive

3.5 Discussion

Despite the technical problems in genotyping offspring in this experiment, it is clear that the races of *M. zebra* do not always mate assortatively under experimental

conditions. It is possible that females from both the Nkhata Bay and Ruarwe races produced hybrid broods, but the only brood shown to include hybrids was from a Ruarwe female in Rep 2. The presence of hybrids in other broods is possible but paternity testing was inconclusive due to PCR problems. Possible hybrids have also been identified in the *M. callainos* experiment, in one brood in each repetition. Again, the paternity of the offspring cannot be ascertained, but hybridisation cannot be ruled out.

It is unfortunate that in Rep 2 of the *M. zebra* experiment, the only broods that amplified were definite or possible hybrids. This prevents an investigation into alternative explanations for the paternity of the broods, apart from that of random mating by the females. It is possible that females would have preferred to mate with the male of their own race, but that the other male was dominant in the experimental tank. Thus the female would not have had free choice of the males present. However, multiple paternity of the brood from female C 995 cannot be ruled out at the present time. If the Ruarwe male did sire some of the offspring in this brood, then the Nkhata Bay male would probably not have been in a position of complete dominance. Multiple paternity has been already demonstrated for field caught mouthbrooding females in Lake Malawi (Parker and Kornfield 1996; Kellogg *et al*, 1997). It has been shown that male dominance can affect female mate choice in laboratory mate choice experiments on African cichlids (Chapter 4, this thesis, review by Turner *et al*, 2001). A further experiment testing the mate choice of females of several geographic races of *M. zebra* where dominance effects are controlled for is currently under way (M.E. Knight, unpubl.). In this design, partitioning experimental tanks with a mesh that prevents the relatively large males from passing through, but allows free movement of the smaller females prevents male-male interactions. It should be noted that male dominance was not apparent in the study of Knight *et al* (1998) where males of three different species were placed in the same tank. If male aggression was a problem in the current study of allopatric populations perhaps this indicates that males recognised each other as possible conspecific competitors.

These preliminary results suggest that populations of the Nkhata Bay and Ruarwe forms of *M. zebra* could interbreed, were it not for their geographic isolation from one another. Although divergence has been very recent (as assumed by their current

conspecific status), each population has developed variations in male colour pattern. According to the results of this study, these variations may not be important to females when choosing a mate, although the majority of the broods appear to be the result of intra-race matings. However an experimental tank is not like the natural environment of the fish. This can be an advantage, for example differences in microhabitat use and seasonality (possible isolating mechanisms) can be disregarded (Knight *et al*, 1998). Tanks are a delimited environment where fish densities and movements may not equate to those under natural conditions and this may result in 'artificial' behaviour. In addition, Carleton *et al* (2000) has recently demonstrated that some Malawi cichlids are capable of detecting UV wavelengths – the illumination provided may not match that of the natural environment. However, tests of assortative mating in sympatric populations (Knight *et al*, 1998) produced results that concurred with field-based differences in population allele frequency distribution (van Oppen *et al*, 1997). This study used the same set up so perhaps environmental conditions in the experimental tanks do not disrupt female mate choice.

Is it possible to determine the specific status of all allopatric populations of mbuna in Lake Malawi? The rapid and ongoing nature of divergence within the lake suggests that one is likely to frequently come across intermediate cases where it is not possible to determine specific status. Indeed, it has been stated by Mayr (1942) that 'every species definition can be only an approach and should be considered with some tolerance' (cited in Coyne *et al*, 1988). Male colour pattern alone may not be enough to isolate populations. Seehausen (1996) suggested that specific status should not be awarded due to variation in male colour pattern alone, but only in cases where at least one other parameter varied as well, for example anatomy or ecology. However, laboratory based experiments of mate choice can highlight the degree of reproductive isolation present, and hint at the possible outcome of environmental changes which would bring two such populations into the same habitat patch.

The tentative results of this experiment indicate a possible lack of reproductive isolation between the *M. zebra* races tested. This would mean that not all allopatric populations, even those differing clearly in male colour, should be regarded as valid species, and estimates for species richness of the Malawi flock should be consequently reduced.

3.5.1 Problems with PCR

A great many offspring generated by this experiment failed to yield a PCR product, or generated products of such low quality that they could not be scored. There are several possible reasons why this happened.

Perhaps the most obvious reason for a lack of PCR product is that the eggs may have been infertile (eggs were not microscopically examined for the presence of an embryo before extraction). It is also possible that very young eggs contained too little DNA to extract properly, given the rather basic extraction protocol. It is known that microsatellite genotyping errors can be caused by extremely low template DNA concentrations (Taberlet *et al*, 1996). It became obvious during the course of this experiment that some loci worked only on certain broods, however, parental genotypes were obtained for each locus tested, ruling out the possibility of null alleles. The loci used appear to have different requirements in terms of minimum DNA concentration or quality.

Another possible reason for PCR failure is that PCR inhibiting compounds were extracted from the samples along with the DNA. One solution to this is to dilute the DNA (and inhibitor) before PCR. Whilst diluting DNA extracted from fin clips generally resulted in amplification, diluting DNA from eggs was rarely successful, possibly due to little of no DNA being present in the extracts. Cleaning the extracts with purification columns was also generally unsuccessful, even when the procedure was used to increase the concentration of the DNA. It is possible that the PCR system used (Reagents, brand of *Taq*, PCR temperature profile) was not correctly optimised for very small quantities of DNA. Parameters were altered in an attempt to increase product yield, but overall, none proved successful. Contamination of PCR reactions by extraneous DNA becomes an important issue when dealing with very low concentrations of template DNA. This may explain some of the cases of non-specific products, or where more than two alleles amplified at a locus.

In order to test whether the PCR system as a whole was at fault, the samples that had been classified as of ambiguous paternity (Table 3.1) were sent to another laboratory (Dr Martin Taylor, University of East Anglia) for further attempts at PCR amplification. Failure of samples to amplify in a second laboratory would indicate that some aspect of the template DNA prevents successful PCR. Current mate choice studies (M.E. Knight, unpublished data) now allow eggs to develop for six days before preservation in order to increase the amount of DNA available for extraction.

3.6 REFERENCES

Albertsen RC and Kocher TD (2001) Assessing morphological differences in an adaptive trait; a landmark based morphometric approach. *Journal of Experimental Zoology* 289: 385-403

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ and Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelliborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London Series B* 266 : 119-130

Bouteillon P (1998) Evolution and Speciation of Cichlid Fishes in Lake Malawi. MPhil thesis, University of East Anglia.

Carleton KL, Harosi FI and Kocher TD (2000) Visual pigments of African cichlid fishes: evidence for ultraviolet vision from microspectrophotometry and DNA sequences. *Vision Research* 40 : 879-890

Coyne JA, Orr HA and Futuyama D (1988) Do we need a new species concept? *Systematic Zoology* 37 : 190-200

Crapon de Caprona MD and Frtzsch B (1984) Interspecific fertile hybrids of haplochromine Cichlidae (Teleostei) and their possible importance for speciation. *Netherlands Journal of Zoology* 43 : 503-538

Danley PD, Markert JA, Arnegard ME and Kocher TD (2001) Divergence with gene flow in the rock dwelling cichlids of Lakw Malawi. *Evolution*

Kellogg KA, Markert JA, Stauffer JR and Kocher TD (1995) Microsatellite variation demonstrates multiple paternity in lekking fishes from Lake Malawi, Africa. *Proceedings of the Royal Society of London Series B* 260 : 79-84

Knight ME and Turner GF (1999) Reproductive isolation among closely related Lake Malawi cichlids : can males recognise conspecific females by visual cues? *Animal Behaviour* 58 : 761-768

Knight ME, Turner GF, Rico C, van Oppen MJH and Hewitt GM (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology* : 1605-1610

Konings A and Dieckhoff HW (1989) *Malawi cichlids in their natural habitat.* Verduijn Cichlids.

Markert JA, Arnegard ME, Danley PD and Kocher TD (1999) Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus :* habitat transience, philopatry and speciation. *Molecular Ecology* 8 : 1013-1026

Mayr E (1942) *Systematics and the Origin of Species.* Columbia University Press, New York.

Parker A and Kornfield I (1996) Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi. *Environmental Biology of Fishes* 47, 345-352

Paterson HEH (1980) A comment on 'mate recognition systems'. *Evolution* 34: 330-331

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC and Sharp BJ (1983) A preliminary survey of the cichlid fishes of rocky habitats of Lake Malawi. *S. Afr. J. Zool.*. 18 : 149-310

Rico C, Kuhnlein J and Fitzgerald GJ (1992) Male reproductive tactics in the threespine stickleback – an evaluation by DNA fingerprinting. *Molecular Ecology* 1 : 79-87

Seehausen O (1996) *Lake Victoria Rock Cichlids*. Verduijn Cichlids, Germany.

Seehausen O, van Alpen JJM (1998) The effect of male colouration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioural Ecology and Sociobiology* 42 : 1-8

Seehausen O, van Alphen JJM and Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277 : 1808-1811

Seehausen O, Witte F, van Alphen JJM and Bouton N (1998) Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *Journal of Fish Biology* 53 (SA) 37-55

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analysis. *Canadian Journal of Zoology* 69 : 82-90

Taberlet P, Griffiths S, Goossens B *et al* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24 : 3189-3194

Turner GF (1999) Explosive Speciation of African Cichlid Fishes. In *Evolution of Biological Diversity* (ed AE Magurran and RM May) Oxford University Press

Turner GF, Seehausen O, Knight ME, Allender CJ and Robinson RL (2001) How many species of cichlid fishes are there in African lakes? *Molecular Ecology*

van Oppen MJH, Rico C, Deutsch JC, Turner GF and Hewitt GM (1997a) Isolation and characterisation of microsatellite loci in the cichlid fish *Pseudotropheus zebra*. *Molecular Ecology* 6 : 387-388

van Oppen MJH, Turner GF, Rico C, Deutsch JC, Ibrahim KM, Robinson RL and Hewitt GM (1997b) Unusually fine scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London Series B* 264: 1803-1812.

van Oppen MJH, Turner GF, Rico C, Robinson RL, Deutsch JC, Genner MJ and Hewitt GM (1998) Assortative mating among rock dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Molecular Ecology* 7: 991-1001

Genotypes for parents and offspring in all repetitions of both experiments. Some alleles deviate from the 2 bp differences in size associated with dinucleotide repeats. This is due to the conversion of the output from the ABI sequencer (given to two decimal places) to a whole number, therefore alleles that differ only by 1 bp should be regarded as the same allele. Locus M11 exhibited both odd and even sized alleles, therefore the 1 bp differences seen at this locus are not an artefact. The rare non-parental alleles seen in a few broods are regarded as a de-novo mutation of the microsatellite or as a PCR or genotyping error. Alleles in italics have originated from an inter-race mating.

Rep 1 *M. callainos* **Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**Pzeb1**	**Pzeb1**	**M11**	**M11**	**UME003**	**UME003**
446	**RP MALE**	**221**	**263**	**112**	**127**	**170**	**250**	**192**	**225**		
450	**NB MALE**	**181**	**227**	**129**	**135**	**160**	**202**	**208**	**227**		
399	**mother brood 62 RP**	**204**	**221**			**162**	**166**				
400A	brood 62	204	263			164	250				
400B	brood 62	204	263			160	168				
400C	brood 62	204	221			164	164				
401	**mother brood 63 RP**	**184**	**217**			**182**	**222**				
697	brood 63	218	221								
698	brood 63										
699	brood 63										
700	brood 63										
420	**mother brood 73 RP**					**178**	**178**				
436	brood 73					170	180				
437	brood 73					170	180				
438	brood 73					170	180				
452	brood 73					170	178				
711	brood 73					178	254				
712	brood 73					170	176				
713	brood 73					178	250				
714	brood 73					170	176				
715	brood 73					176	250				
716	brood 73					176	250				

Rep 1 *M. callainos* **Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**		**Pzeb1**	**Pzeb1**		**M11**	**M11**	**UME003**	**UME003**
446	**RP MALE**	**221**	**263**	**112**	**127**		**170**	**250**		**192**	**225**		
450	**NB MALE**	**181**	**227**	**129**	**135**		**160**	**202**		**208**	**227**		
318B	**mother brood 53 NB**						**182**	**208**					
319	brood 53						160	182					
661	brood 53						182	200					
662	brood 53						182	202					
663	brood 53						160	182					
664	brood 53						160	208					
421	**mother brood 74 NB**	**221**		**115**	**129**		**160**	**176**		**204**	**220**		
422A	brood 74												
422B	brood 74	221	227	115	129		160	176		220	227		
422C	brood 74						160	160		*192*	??MESS		
422D	brood 74												
422E	brood 74	192	221	*112*	*127*	139	*168*	176		*192*	211		
423A	brood 74	221	227	115	129		158	176		208	220		
423B	brood 74			115	*127*	135	160	176					
423C	brood 74						164	176					
423D	brood 74						162	176					
423E	brood 74												
424A	brood 74						174	160					
424B	brood 74			115			162	176					
424C	brood 74			129	131		160	160		208	220		
447	**mother brood 77 NB**	**231**	**237**	**127**	**152**		**152**	**180**					
448A	brood 77												
448B	brood 77												
448C	brood 77												
448D	brood 77	227	237	129	152		158	180					
448E	brood 77												
449A	brood 77												
449B	brood 77						178	162	20?				
449C	brood 77			129.39									
449D	brood 77												

Rep 1 *M. callainos* **Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**Pzeb1**	**Pzeb1**	**M11**	**M11**	**UME003**	**UME003**
446	**RP MALE**	**221**	**263**	**112**	**127**	**170**	**250**	**192**	**225**		
450	**NB MALE**	**181**	**227**	**129**	**135**	**160**	**202**	**208**	**227**		
449E	brood 77					150	200				
449F	brood 77										

M. callainos **Rep 2** **Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**PZEB1**	**PZEB1**		**M11**	**M11**
434	**RP MALE**	**184**	**196**	**123**	**141**	**146**	**176**		**210**	**223**
435	**NB MALE**	**209**	**225**	**127**	**148**	**184**	**200**		**199**	**204**
470	**mother brood 87 RP**	**221**	**225**			**188**	**202**			
717	brood 87	196	221							
718	brood 87	196	225			186	142	176		
719	brood 87	184	225			202	146			
720	brood 87	215	225			174	202			
721	brood 87	184	221			146	188			
722	brood 87									
723	brood 87	184	221			146	202			
724	brood 87	196	221			174	188			
725	brood 87	196	221			174	202			
726	brood 87					174	188			
727	brood 87	196	225			176	188			
728	brood 87	196	225			174	202			
729	brood 87	184	221			174	188			
730	brood 87	184	221			174	202			
731	brood 87					174	202			
732	brood 87	184	225			176	202			

M. callainos Rep 2 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	PZEB1	PZEB1		M11	M11	
434	**RP MALE**	**184**	**196**	**123**	**141**	**146**	**176**		**210**	**223**	
435	**NB MALE**	**209**	**225**	**127**	**148**	**184**	**200**		**199**	**204**	
564	**mother brood 97 RP**	**221**	**225**			**188**	**204**				
565A	brood 97					*184*	142	176			
565B	brood 97										
565C	brood 97					142	184				
566A	brood 97										
566B	brood 97					142	186				
566C	brood 97					142	186				
567A	brood 97					176	202				
567B	brood 97	196	221			174	202				
567C	brood 97	184	221			174	202				
568A	brood 97										
568B	brood 97	184	225			146	202				
568C	brood 97					144	202				
569A	brood 97	196	221			174	202				
569B	brood 97					146	188				
569C	brood 97					174	188				
459	**mother brood 81 NB**	**195**	**221**	**129**	**148**	**174**	**222**		**?**	**?**	
460A	brood 81								210	229	
460B	brood 81			127	148						
460C	brood 81			127	129	222	184		204	223	
460D	brood 81			127	129	184	222		204	227	
460E	brood 81			127	129	*174*	*174*				
461A	brood 81					*174*	*174*		204	223	
461B	brood 81										
461C	brood 81								204	227	
461D	brood 81					*174*	222		199	213	223
461E	brood 81			129	148	*174*	222		199	223	
462A	brood 81			127	148						
462B	brood 81	194	225	127	129	*174*	222		204	223	
462C	brood 81					*174*	*174*				
462D	brood 81			127	129	*174*	*174*		199	213	223

M. callainos Rep 2 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	PZEB1	PZEB1	M11	M11	
434	**RP MALE**	**184**	**196**	**123**	**141**	**146**	**176**	**210**	**223**	
435	**NB MALE**	**209**	**225**	**127**	**148**	**184**	**200**	**199**	**204**	
505	**mother brood 88 NB**	**203**	**221**			**188**	**206**			
507	brood 88	209	221			186	186			
508	brood 88	221	225			184	184			
509	brood 88	204	209			188	200			
750	**mother brood 112 NB**	**176**	**231**							
751	brood 112	225	231							
752	brood 112	209	231							
753	brood 112	209	231							
754	brood 112	176	209							
755	brood 112	209	231							
756	brood 112									
757	brood 112	176	225							
767	**mother brood 117 NB**	**176**	**231**							
768	brood 117	216	231							
769	brood 117	176	225							
770	brood 117	225	231							
771	brood 117	225	231							
772	brood 117	176	225							
773	brood 117	216	231							
774	brood 117	225	231							
775	brood 117	176	215							
776	brood 117									
1292	**mother brood 142 NB**	**192**	**262**	**121**	**157**			**193**	**229**	**231**
1293	brood 142							192	231	
1294	brood 142									
1295	brood 142									
1296	brood 142									
1297	brood 142									
1298	brood 142									
1299	brood 142									
1300	brood 142									

M. callainos Rep 2 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	PZEB1	PZEB1	M11	M11
434	**RP MALE**	**184**	**196**	**123**	**141**	**146**	**176**	**210**	**223**
435	**NB MALE**	**209**	**225**	**127**	**148**	**184**	**200**	**199**	**204**
1301	brood 142								
1302	brood 142								
1303	brood 142								

M. zebra Rep 1 Sample No.	Sample ID	UNH130	UNH130	DXT3	DXT3	PZEB1	PZEB1	
162	**MALE C**	**185**	**241**	**139**	**148**	**144**	**164**	
163	**MALE NB**	**217**	**243**	**133**	**137**	**182**	**204**	
313	**mother brood 51 C**					**172**	**216**	
314A	brood 51					144	172	
314B	brood 51					144	216	
314C	brood 51					144	215	
314D	brood 51					164	216	
315A	brood 51					164	172	
315B	brood 51					144	172	
315C	brood 51					164	172	
315D	brood 51					144	216	
316A	brood 51					164	172	
316B	brood 51					164	216	
316C	brood 51					164	172	
316D	brood 51					144	216	
317A	brood 51					164	172	
317B	brood 51					144	172	
317C	brood 51					164	216	
317D	brood 51							
542	**mother brood 94 C**	**186**	**233**			**144**	**162**	
543B	brood 94					144	144	
543C	brood 94					144	144	164
543D	brood 94					162	162	
543E	brood 94					142	162	
543F	brood 94	233	241			144	162	
544A	brood 94	184	233			144	144	
544B	brood 94	184	186			144	144	
544C	brood 94	184	186			144	162	
544D	brood 94					144	144	
544E	brood 94					144	162	
544F	brood 94					142	164	
545A	brood 94					162	162	
545B	brood 94					144	16	
545C	brood 94	233	241			144	162	184
545D	brood 94					144	144	
545E	brood 94					144	162	
545F	brood 94	233	241			144	164	
546A	brood 94					142	164	
546B	brood 94	233	241			144	12	
546C	brood 94					144	144	164
546D	brood 94					142	142	
546E	brood 94	186	241			144	162	
546F	brood 94	186				144	164	
547A	brood 94	186				164	164	
547B	brood 94	184				144	144	
547C	brood 94	184	233			162	162	
547D	brood 94	186	241			162	162	
547E	brood 94	184	233			144	164	
547F	brood 94	186	241			144	162	
548A	brood 94					142	162	
548B	brood 94					142	142	164
548C	brood 94	186	241			162	144	

M. zebra **Rep 1** **Sample No.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**PZEB1**	**PZEB1**	
162	**MALE C**	**185**	**241**	**139**	**148**	**144**	**164**	
163	**MALE NB**	**217**	**243**	**133**	**137**	**182**	**204**	
548D	brood 94	184	186			144	144	164
548E	brood 94	186				142	164	
548F	brood 94	184	186			144	160	
549A	brood 94	233				144	164	
549B	brood 94	184				162	162	
549C	brood 94					144	162	
549D	brood 94					144	144	
549E	brood 94	186	241			162	162	
549F	brood 94					142	142	
550A	brood 94	184				162	162	
550B	brood 94					142	164	
550C	brood 94					142	164	
550D	brood 94	184	233			144	144	
550E	brood 94					144	164	
550F	brood 94	186	241			142	164	
551A	brood 94	186	241			164	162	
551B	brood 94					144	164	
551C	brood 94					144	164	
551D	brood 94					144	162	
551E	brood 94					142	164	*184*
551F	brood 94					164	164	
552A	brood 94	233	241			144	12	
552B	brood 94					144	144	
552C	brood 94	233	241			144	144	184
552D	brood 94					144	144	
318A	**mother brood 52 C**	**185**	**209**			**162**	**194**	
651	brood 52					144	14	
652	brood 52					162	162	
653	brood 52					162	162	
654	brood 52	208	241			144	162	
655	brood 52					144	162	194
656	brood 52					144	196	
657	brood 52	184				162	142	*200*
658	brood 52	208	241			144	192	
659	brood 52	184	241			144	194	
660	brood 52	184				144	164	
406	**mother brood 66 NB**					**174**	**190**	
407A	brood 66					174	166	
407B	brood 66							
407C	brood 66							
407D	brood 66					176	164	
407E	brood 66					174	144	
408A	brood 66					144	174	
408B	brood 66					174	164	
408C	brood 66							
408D	brood 66							
408E	brood 66							
469	**mother brood 86 NB**	**209**	**227**	**137**	**133**	**142**	**166**	
471A	brood 86	209		133	137	142	202	
471B	brood 86	227		133	137	166	182	
471C	brood 86					-		

M. zebra **Rep 1** **Sample No.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**PZEB1**	**PZEB1**	
162	**MALE C**	**185**	**241**	**139**	**148**	**144**	**164**	
163	**MALE NB**	**217**	**243**	**133**	**137**	**182**	**204**	
472A	brood 86	227		133	137	168	184	
472B	brood 86	209		137		142	182	
472C	brood 86	209		137	133	168	204	
473A	brood 86					-		
473B	brood 86	227		133	137	142	204	
473C	brood 86	227		133	137	142	182	
474A	brood 86	209		133	137	166	204	
474B	brood 86					-		
474C	brood 86	227		133		142	182	
475A	brood 86	209		133	137	142	204	
475B	brood 86	227		133	137	142	204	
474C	brood 86	209		133	137	142	204	
476A	brood 86	227		133		142	182	
476B	brood 86	209		133	137	168	182	
476C	brood 86					142	182	
476D	brood 86	227		137		168	204	
536	**mother brood 93 NB**					**166**	**168**	
537A	brood 93					*142*	168	
537B	brood 93							
537C	brood 93							
537D	brood 93					164	202	
537E	brood 93					168	*142*	204
538A	brood 93					164	*142*	
538B	brood 93					168	146	178
538C	brood 93					*142*	164	204
538D	brood 93					168	176	
538E	brood 93							
539B	brood 93							
539C	brood 93							
539D	brood 93					*144*	164	
539E	brood 93							
540B	brood 93					166	180	
540C	brood 93							
540D	brood 93					146	166	
540E	brood 93							
541B	brood 93					*140*	166	202
541C	brood 93					*146*	164	
541D	brood 93					*146*	166	204
553	**mother brood 95 NB**	**185**	**215**					
556	brood 95	185	243					
557	brood 95	185	217					
558	brood 95	215	243					
559	brood 95	215	243					

M. zebra Rep 2 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	UME003	UME003	M11	M11
783	**MALE C**	**184**	**233**	**121**	**131**	**222**	**220**	**221**	
784	**MALE NB**	**202**	**218**	**129**	**131**	**226**	**230**	**221**	**225**
785	**mother brood 118 C**	**208**	**214**						
786	brood 118								
787	brood 118								
788	brood 118								
789	brood 118								
790	brood 118								
791	brood 118								
792	brood 118								
793	brood 118								
794	brood 118								
795	brood 118								
796	brood 118								
797	brood 118								
798	brood 118								
799	brood 118								
800	brood 118								
801	brood 118								
802	**mother brood 119 NB**	**220**	**232**			**210**	**238**		
803	brood 119								
804	brood 119								
805	brood 119								
806	brood 119								
807	brood 119								
808	brood 119								
809	brood 119								
810	brood 119								
811	brood 119								
812	brood 119								
813	brood 119								
814	brood 119								
815	brood 119								
816	brood 119								
817	brood 119								
818	brood 119								
819	brood 119								
820	brood 119								
821	brood 119								
822	brood 119								
823	brood 119								
824	brood 119								
825	**mother brood 120 NB**	**210**	**216**			**218**	**230**		
826	brood 120								
827	brood 120								
828	brood 120								
829	brood 120								
830	brood 120								
831	brood 120								
832	brood 120								

***M. zebra* Rep 2 Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**UME003**	**UME003**	**M11**	**M11**
783	**MALE C**	**184**	**233**	**121**	**131**	**222**	**220**	**221**	
784	**MALE NB**	**202**	**218**	**129**	**131**	**226**	**230**	**221**	**225**
833	brood 120								
834	brood 120								
835	brood 120								
836	brood 120								
837	brood 120								
838	brood 120								
839	brood 120								
840	brood 120								
841	brood 120								
842	brood 120								
882	**mother brood 123 NB**	**216**	**228**						
883	brood 123								
884	brood 123								
885	brood 123								
886	brood 123								
887	brood 123								
888	brood 123								
889	brood 123								
890	brood 123								
891	brood 123								
892	brood 123								
893	brood 123								
894	brood 123								
895	brood 123								
896	brood 123								
897	brood 123								
898	brood 123								
899	brood 123								
900	brood 123								
901	brood 123								
902	brood 123								
903	brood 123								
904	brood 123								
905	**mother brood 124 C**	**214**	**234**			**153**	**234**		
906	brood 124								
907	brood 124	*202*	214			238	236		
908	brood 124	*202*	234			234			
909	brood 124	*202*	234			234	238		
910	brood 124					234			
911	brood 124					234	238		
912	brood 124					234	153		
913	brood 124					234	238		
914	brood 124					234			
952	**mother brood 127 NB**	**218**	**220**			**220**	**230**		
953	brood 127								
954	brood 127								
955	brood 127								
956	brood 127								
957	brood 127								
958	brood 127								

M. zebra **Rep 2 Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**UME003**	**UME003**	**M11**	**M11**
783	**MALE C**	**184**	**233**	**121**	**131**	**222**	**220**	**221**	
784	**MALE NB**	**202**	**218**	**129**	**131**	**226**	**230**	**221**	**225**
959	brood 127								
960	brood 127								
961	brood 127								
962	brood 127								
963	brood 127								
964	brood 127								
965	brood 127								
966	brood 127								
967	brood 127								
968	brood 127								
969	brood 127								
970	brood 127								
971	brood 127								
972	brood 127								
973	brood 127								
974	brood 127								
975	brood 127								
976	brood 127								
977	brood 127								
995	**mother brood 129 C**	**184**	**234**	**125**	**131**				
996	brood 129	*202*	234						
997	brood 129	184	208	129	131				
998	brood 129	*202*	234	125	131				
999	brood 129	184							
1000	brood 129	234		131					
1001	brood 129	*202*	234	125	129				
1002	brood 129	184	*218*	125	131				
1003	brood 129	184	*218*	125	129				
1004	brood 129	*217*	234	125	129				
1005	brood 129	*202*	234	125	129				
1006	brood 129	184	*202*	125	131				
1007	brood 129	*202*	234	129	131				
1008	brood 129	*202*	234	129	131				
1009	brood 129	*202*	234	131					
1010	brood 129	*202*	234	125	129				
1011	brood 129	*202*	234	131					
1012	brood 129	184		125	131				
1013	brood 129			129	131				
1014	brood 129	184	*218*	125	129				
1015	brood 129	184	*202*	125	129				
1016	brood 129	*202*	234	125	129				
1017	brood 129	184		129	131				
1018	brood 129			125	129				
1019	brood 129	234		131					
1020	brood 129	202	234	125	129				
1021	**mother brood 130 NB**	**202**	**212**	**129**	**??**	**226**	**244**	**219**	**221**
1022	brood 130			125	129	228			
1023	brood 130			132		208	220		
1024	brood 130								
1025	brood 130			113	129				

M. zebra Rep 2 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	UME003	UME003	M11	M11
783	**MALE C**	**184**	**233**	**121**	**131**	**222**	**220**	**221**	
784	**MALE NB**	**202**	**218**	**129**	**131**	**226**	**230**	**221**	**225**
1026	brood 130					206			
1027	brood 130								
1028	brood 130								
1029	brood 130								
1030	brood 130					234	238	208	221
1031	brood 130								
1032	brood 130								
1033	brood 130								
1034	brood 130								
1035	brood 130								

M. zebra REP3 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	UME003	UME003	M11	M11
781	**MALE C**	**214**	**233**	**121**	**146**	**222**	**220**	**221**	
782	**MALE NB**	**211**	**219**	**129**	**131**	**226**		**209**	**229**
843	**mother brood 121 C**	**220**	**232**			?	?		
844	brood 121	207							
845	brood 121								
846	brood 121	207	214						
847	brood 121	207	214						
848	brood 121								
849	brood 121								
850	brood 121								
851	brood 121								
852	brood 121								
853	brood 121								
854	**mother brood 122 C**	**184**	**186**	**125**	**131**	**222**	**144**		
855	brood 122			121	131	222	144		
856	brood 122			125	146	222			
857	brood 122					222			
858	brood 122								
859	brood 122					220	144		
860	brood 122					222			
861	brood 122					222			
862	brood 122								
863	brood 122					220	222		
864	brood 122					222			
865	brood 122					222	144		
866	brood 122					222	144		
867	brood 122								
868	brood 122								
869	brood 122					222			
870	brood 122					222	220		
871	brood 122					222	204		
872	brood 122								

M. zebra REP3 Sample no.	Sample ID	UNH130	UNH130	DXT3	DXT3	UME003	UME003	M11	M11
781	**MALE C**	**214**	**233**	**121**	**146**	**222**	**220**	**221**	
782	**MALE NB**	**211**	**219**	**129**	**131**	**226**		**209**	**229**
873	brood 122								
874	brood 122								
875	brood 122					222			
876	brood 122					220	144		
877	brood 122								
878	brood 122					222			
879	brood 122								
880	brood 122					222	144		
881	brood 122					220	144		
915	**mother brood 125 C**	**184**	**186**	**125**	**146**				
916	brood 125	184	214						
917	brood 125	186	214						
918	brood 125	186	214						
919	brood 125	186	233						
920	brood 125	184	213						
921A	brood 125	186	214						
921B	brood 125	186	233f						
922	brood 125	186	214						
923	brood 125	186	233						
924	brood 125								
925	brood 125	184	233	121	146				
926	brood 125	184	211	121	146				
927	brood 125	186	233						
928	brood 125	186	214						
929	brood 125	184	214						
930	brood 125	186	214						
931	brood 125	186	214						
932	brood 125	184	214						
933	brood 125	184	233						
934	brood 125	184	233						
935	brood 125	184	214						
936	brood 125	186	233						
937	brood 125	186	214						
938	brood 125	184	214						
939	brood 125	184	214						
940	brood 125	186	214						
941	brood 125	186	214						
942	brood 125	186	214						
943	brood 125	186	214						
944	brood 125	184							
945	brood 125								
946	**mother brood 126 NB**	**202**		**133**	**141**				
947	brood 126	202	211	131	133				
948	brood 126	202	219	131	133				
949	brood 126	202	219	129	133				
950	brood 126			131	141				
951	brood 126			131	141				
978	**mother brood 128 NB**	?	?			**208**	**224**	**208**	**221**

M. zebra **REP3 Sample no.**	**Sample ID**	**UNH130**	**UNH130**	**DXT3**	**DXT3**	**UME003**	**UME003**	**M11**	**M11**
781	**MALE C**	**214**	**233**	**121**	**146**	**222**	**220**	**221**	
782	**MALE NB**	**211**	**219**	**129**	**131**	**226**		**209**	**229**
979	brood 128					208	226	229	221
980	brood 128					224	226	209	
981	brood 128					224	226		
982	brood 128					224	226	209	221
983	brood 128					220	226	221	
984	brood 128					220	224	204	208
985	brood 128	219				224	226	209	
986	brood 128					224	226	209	229
987	brood 128							209	223
988	brood 128								
989	brood 128								
990	brood 128								
991	brood 128								
992	brood 128								
993	brood 128								
994	brood 128								
1036	**mother brood 131 NB**	**202**				**222**			
1037	brood 131	202				220	235		
1038	brood 131	202				220	226		
1039	brood 131					226	232		
1040	brood 131					220	226		
1041	*brood 131*								
1042	brood 131								
1043	brood 131								
1044	brood 131								
1045	brood 131					226	235		
1046	brood 131								
1047	brood 131					226			
1048	brood 131					235			
1049	brood 131					226	235		

4 Mate Choice of Cichlid Fish from Lake Victoria

4.1 Chapter Note

This chapter is the result of joint research by C.J. Allender and O. Seehausen. O. Seehausen carried out experimental design, fish maintenance and behavioural observations together with Martine Maan and Robert Fraser. C.J. Allender carried out DNA extraction and microsatellite genotyping.

4.2 Introduction

4.2.1 Lake Victoria

Lake Victoria (Figure 4.1) is one of the three Great Lakes of Eastern Africa. It is the third largest lake on Earth, and was formed by uplift of the area to the west of the current lake, which caused ponding of the westward flowing rivers in the region. This is in contrast to lakes Malawi and Tanganyika that were formed by water accumulation in deep rift valleys and are 785m and 1470m in depth respectively. Lake Victoria is relatively shallow with a maximum depth of 70m and has a surface area of 70 000 km^2 (Martens, 1997). The large surface area and shallow depth of the lake allow complete mixing of the water column, unlike Lakes Malawi and Tanganyika, and this permits fish to occupy any depth habitat within the lake.

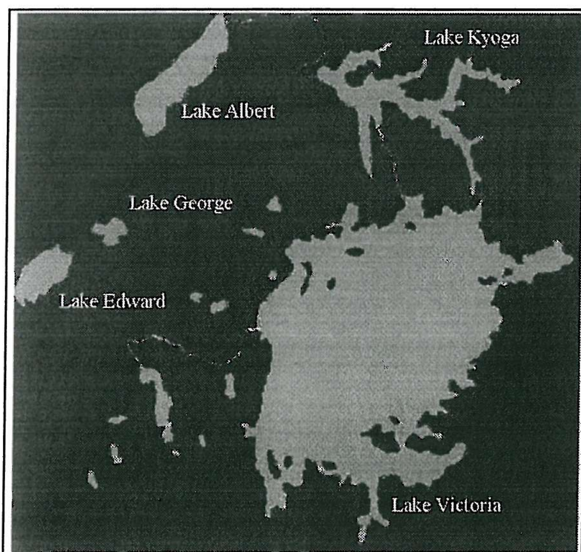


Figure 4.1. Lake Victoria and surrounding lakes. Lake Victoria has a maximum width of 320 km and a maximum length of 400 km. The relative locations of Lakes Victoria, Malawi and Tanganyika may be seen in Figure 1.2.

4.2.2 The cichlids of Lake Victoria

Lake Victoria is also home to a large radiation of cichlids. Seehausen *et al* (1997b) estimate the number of species present as >500. The term 'super species flock' has been used in the case of Lake Victoria due to the inclusion of species found in surrounding lakes and rivers (Greenwood, 1984). Reconstruction of the phylogeny of this flock has proved difficult due to a lack of variation in the markers surveyed, for example Sage *et al* (1984) found interspecific differences of just 0.006 detectable substitutions per locus in a survey of 10 allozyme loci. DNA sequencing has provided more insight; Booton *et al* (1999) suggested a polyphyletic origin of the Lake Victoria flock based on sequence data from the ribosomal RNA internal transcribed spacer one. In this study, phylogenetic analysis resulted in a sequence from Lake Edward clustering within the Lake Victoria sequences, and the authors attributed this to 'reinvasion by the products of earlier cladogenesis events'. These results contrast those of Meyer *et al* (1990) who suggested a monophyletic origin for the species flock within Lake Victoria. However, their conclusions were based on a limited number of samples with none from other lakes in the region. In a more thorough investigation, Nagl *et al* (2000) proposed a more complex scenario as they detected seven mitochondrial DNA haplotypes at a regional level. Endemic haplochromine cichlids from Lake Victoria fell into a single haplogroup, but this grouping was not a monophyletic assemblage and also included species found in other lakes and rivers. However, a large proportion of Lake Victoria endemics clustered into one subgroup within this haplogroup. The authors concluded that while a significant radiation took place within the lake from a single founding population, the contemporary flock is the result of the divergence of several ancestral lineages. This may also explain the fact that molecular evidence dates the radiation of the modern flock at around 200000 years ago (Meyer *et al*, 1990), geological evidence suggests modern Lake Victoria was dry as little as 15000 years ago (Johnson *et al*, 1996).

The species flock of Lake Victoria also contains a group of rock-dwelling cichlids similar in ecology to the mbuna of Lake Malawi. This group of fish is described by the local name 'mbipi' and the species composition and ecology of the mbipi have been described in detail by Seehausen (1996). Like the mbuna of Lake Malawi, the mbipi are sexually dimorphic maternal mouthbrooders, and they exhibit geographical variation in

male colour patterns, and polymorphisms are often present within a population. As in Lake Malawi, given the enormous diversity in both colour patterns and ecology in cichlids from Lake Victoria, questions must be asked regarding mechanisms of speciation and indeed what constitutes a species in this group. Post-zygotic isolation appears to be limited as viable hybrids can be generated in the laboratory (Seehausen *et al*, 1997a). Seehausen and van Alphen (1999) present evidence for the occurrence of both allopatric and sympatric speciation through a plot of numbers of closely related species pairs against amount of geographic range overlap. A bimodal distribution was detected with the largest numbers of species pairs existing in almost complete sympatry or allopatry, with relatively few pairs having intermediate levels of range overlap. A possible mechanism of sympatric speciation was indicated as disruptive sexual selection on male colour patterns by the authors as species pairs which exist in sympatry differ more than allopatric pairs in male colour (the trait under selection) in the way that has been shown to affect female preference (Seehausen and van Alphen, 1998; Seehausen *et al*, 1999). The specific status of colour morphs within populations is not often obvious. Seehausen *et al* (1999) investigated the genetics of mate preference and colour patterns in different colour morphs of *Neochromis omnicaeruleus*. They concluded that the morphs represented one original and two incipient species with reduced gene flow between them due to mate preference and highly distorted sex ratios in broods from certain crosses.

The approach of comparing ecological observations with behavioural experiments in the laboratory appears to be a useful approach for testing reproductive isolation between colour morphs. Field sampling allows assessment of potential differences in habitat use, food intake and reproductive timing (Seehausen *et al*, 1999). Microsatellite typing of population samples may reveal isolation implied by differences in allele frequencies (for example van Oppen *et al*, 1998). Carefully designed laboratory experiments on mate choice may reveal the presence of pre-zygotic barriers to gene flow between colour morphs (Knight *et al*, 1998; Seehausen and van Alphen, 1998; Seehausen *et al*, 1999 and reviewed by Turner *et al*, 2001). In addition, it may be possible to generate hybrids of two different colour morphs in order to examine mate preference and viability (Seehausen, unpublished).

4.2.3 Aims

Pundamilia pundamilia (Seehausen and Bouton, 1998) and *P. nyererei* (Witte and Witte-Mass, 1985) represent a sympatric species pair of rock dwelling cichlids from Lake Victoria. They are very similar in appearance apart from the dorsal area of mature sexually active males (red in *P. nyererei* and blue in *P. pundamilia* – see Figure 4.2). As their Latin names imply, they have already been formally described as separate species although Seehausen and van Alphen (1998) note that in some parts of their range they behave as colour morphs of the same species and in others as separate species. Hybrids between the two species are easily generated in the laboratory but are not normally observed in the natural environment except in areas of greatly reduced water clarity (Seehausen *et al*, 1997a). The aim of this study is to test for reproductive isolation between the two species using microsatellite based paternity assignment of the offspring generated from four mate choice trials. The utility of microsatellite loci originally isolated from species from Lake Malawi and elsewhere is assessed for this purpose. In addition, partial microsatellite typing of broods from F1 hybrids of these two species is used as an indication of mate preference in order to compare real mating patterns of F1 hybrid females with behavioural indicators of hybrid mate preference. This allows investigation of the heritable basis of female preference, and also permits the assumptions of a model of sympatric speciation by sexual selection to be tested, namely that female preference is a single locus trait with complete dominance (as in the model of Turner and Burrows, 1995).

4.3 Methods

4.3.1 Experimental set-up

Two different designs were used during the course of this experiment. A number of *P. pundamilia*, *P. nyererei* and hybrid females were presented with a choice of three males (one from each species and one hybrid). The design was varied in order to allow or prevent interaction between the males. This was achieved by partitioning the experimental tank into three sections with a mesh which prevented the large males from moving out of their respective section, but which allowed the smaller females unrestricted access to any male. Removing the mesh partitions (an ‘open’ design)

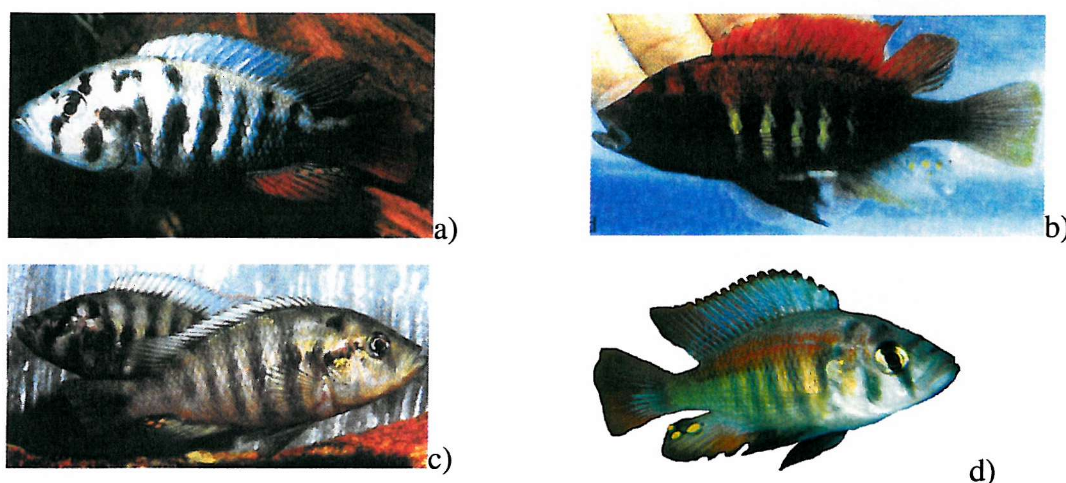


Figure 4.2. Representatives of the species used in this study. a) *P. pundamilia* male, b) *P. nyererei* male. The contrast between the dorsal areas of these fish is immediately apparent. c) *P. pundamilia* females. *P. nyererei* females are very similar to *P. pundamilia* apart from subtle differences in head anatomy. d) Male hybrid between *P. pundamilia* and *P. nyererei*. Photos taken by O. Seehausen, reproduced with kind permission.

permitted male interaction. The experiment was subdivided into four trials, details of which may be found in Table 4.1. Females were allowed to spawn freely, and then eggs were removed after 3 to 7 days of mouthbrooding. Fin clips from the adult fish and eggs or fry were preserved in absolute ethanol. The standard length of each male was measured and an index of coloration was recorded as follows. The hue and intensity of pigmentation of the dorsum of each male was assigned a score of -1 to 3, -1 representing blue, 0 dull, 1 slightly red, 2 red, 3 intensely red. The colour score recorded represents the mean of scores assigned to each fish by at least two researchers.

The experiment consisted of four separate trials. Each tested assortative mating between the females of the different species and also examined hybrid mate preference. Some males and females were used twice in different trials (see Table 4.1). Absolute mate choice of all *P. pundamilia* and *P. nyererei* females was assessed by microsatellite genotyping their entire brood. A preliminary investigation of the mate preference of F_1 hybrid females was carried out by partially typing all of the broods resulting from one trial (Trial 2.1) of the experiment.

4.3.2 Genetic Analysis

Samples were washed thoroughly with Millipore water before a standard proteinase K/chloroform extraction protocol was carried out (Rico *et al*, 1992, see Chapter 2). Prior to extraction, as much of the yolk sac as possible was removed from the larva. Young eggs with no visible larvae were extracted in their entirety. DNA was resuspended in 200 μL Tris-HCl buffer (10mM) apart from that from very young eggs which was resuspended in 100μL of the same buffer.

The PCR protocols and procedures described in Chapter 2 were followed. Five microsatellite loci previously optimised for Malawi cichlids were tested on two broods to determine which were the most polymorphic in these Victoria species; UNH130, UME003, DXTUCA3, UNH106 and M11 (see Table 2.2 for locus characteristics and original references). Two loci (UNH106 and DXTUCA-3) were not variable enough amongst the males tested to provide good assignments of paternity (data not shown). The remaining three loci exhibited high levels of variation and were thus used to type the rest of the broods in the experiment.

An assay of the amount of genetic variation present in the laboratory populations of *P. pundamilia* and *P. nyererei* was carried out by calculating allele frequency and heterozygosity for the three loci used in this study. Genotype data for all the adults from the mate choice experiments was included along with data collected for 10 males not involved in the trials. Individuals that had not been unambiguously identified during the trials were excluded from these calculations (there were two cases of dubious identification of females as either *P. pundamilia*, *P. nyererei* or F1 hybrid as they are very similar in appearance).

In an attempt to determine if it was necessary to type every individual in each brood to obtain an accurate estimate of mate preference, the BROOD program (Dewoody *et al*, 2000) was used to generate one hundred simulated broods based on the combined allele frequencies of the *P. pundamilia* and *P. nyererei* populations. The genotypes of the parents for each brood were generated randomly based on the observed allele frequencies, assuming Hardy-Weinberg equilibrium. The offspring genotypes resulted from Mendelian inheritance of parental alleles. Broods were simulated with a

maximum of three sires, as this was the number of males present in each experimental tank. Each brood was resampled one hundred times to determine the number of offspring that had to be typed before all parental alleles were recovered (including those that are identical in state). The mean across all broods gives the parameter N*, and due to the number of broods generated, 95% confidence intervals of this parameter may also be calculated. This will provide an indication of the number of offspring that should be typed from each brood to be reasonably certain that all paternal alleles have been detected. Broods of 40 and 50 offspring were simulated, assuming an equal paternal contribution of three males. In addition, the number of males and their relative contributions were varied whilst holding brood size constant at 50 offspring.

Trial	Female ID	N in brood	Males Present									Design	Notes
				SL	CS		SL	CS		SL	CS		
1.1	B1	34	B1	78	-1	R1	77	3	H1	77	1	O	
	B2	4	B1	78	-1	R1	77	3	H1	77	1	O	
1.2	B1	31	B7s	91	-1	R4s	80	1.5				P	
	R1	19	B7s	91	-1	R3s	84	3				P	
	R2	25	B7s	91	-1	R3s	84	3				P	
	R9	1**	B7s	91	-1	R3s	84	3				P	
2.1	2B1	43	B1	78	-1	R1	77	3	H2	78	1	O	
	2B2	50	B2	68	-1	R1	77	3	H2	78	1	O	R1 Dominant
	2B3	48	B4	88	-1	R1	77	3	H2	78	1	P	
	2B4	65	B4	88	-1	R1	77	3	H2	78	1	P	
	2R?2	50	B4	88	-1	R1	77	3	H2	78	1	P	
2.2	2B	30	B1s	79	-1	R2s	85	3				P	ID Problem
	2R?2	34	B1s	79	-1	R2s	85	3				P	

Table 4.1. Details of the four trials of the experiment. Species ID as follows ; B = *P. pundamilia*, R = *P. nyererei*, H = hybrid. An ID problem is noted for female 2B in trial 2.2 as two fin clips were provided with the brood which were revealed as coming from different individuals after genotyping. The clip with the genotype compatible with the offspring was assumed to have come from the mother of this brood. O denotes an open design, P a partitioned design. ** the majority of the brood from female R9 in Trial 1.2 was lost prior to preservation apart from a single individual. SL – standard length of male (mm). CS – colour score (see text for description). Only females that spawned are listed.

Trial	Female ID	N in brood	Males Present									Design	Notes
				SL	CS		SL	CS		SL	CS		
2.1	2H1	14	B1	78	-1	R1	77	3	H2	78	1	O	
	2H2	26	B3	~90	?	R1	77	3	H2	78	1	O	
	2H3	11	B3	~90	?	R1	77	3	H2	78	1	P	
	2H4	17	B4	88	-1	R1	77	3	H2	78	1	P	
	2H5	32	B4	88	-1	R1	77	3	H2	78	1	P	
	2H6	34	B4	88	-1	R1	77	3	H2	78	1	P	ID problem
	2H7	28	B4	88	-1	R1	77	3	H2	78	1	P	
	2H9	18	B4	88	-1	R1	77	3	H2	78	1	P	
	2H11	33	B4	88	-1	R1	77	3	H2	78	1	P	Preservation problem?
	2H12	28	B4	88	-1	R1	77	3	H2	78	1	P	

Table 4.2. **Number of females and brood sizes of the F1 hybrid females from Trial 2.1. Species identification as Table 4.1. The brood of female 2H11 was noted to be particularly poorly preserved upon extraction. Female 2H6 was thought to be a misidentified *P. nyererei*, but comparison of the three locus genotype for this female with the genotypes of known *P. pundamilia* and *P. nyererei* individuals revealed alleles present in both species. This female was therefore reassigned as a hybrid. O denotes and open experimental design, P a partitioned design. SL – standard length of male (mm), CS – colour score(see text for description). ? indicates colour score was not recorded.**

4.4 Results

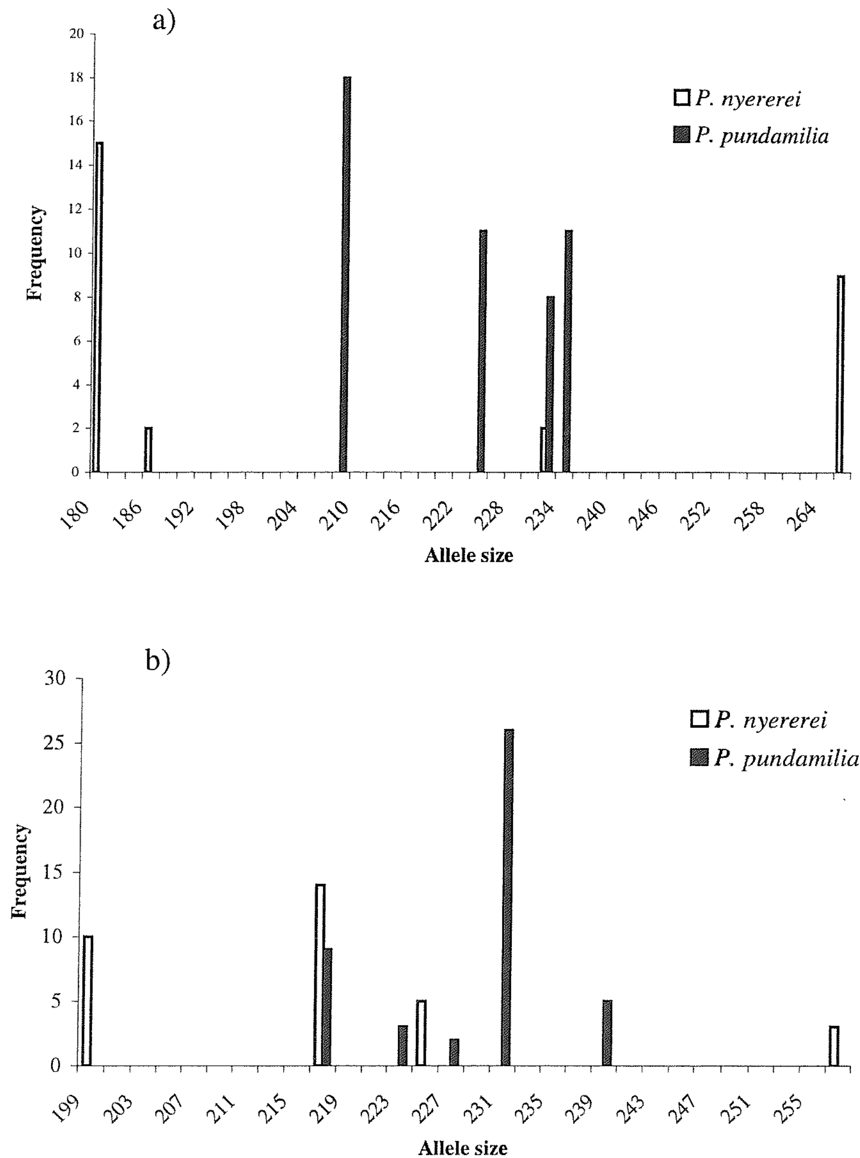
4.4.1 Utility of Loci chosen

The loci chosen worked relatively well in the species from Lake Victoria used in this study, despite the fact that they were originally isolated from *Oreochromis niloticus*, *Metriaclima zebra* and the Tanganyikan cichlid *Tropheus moorii*. However, the stutter pattern associated with UME003 alleles was noted to be less easy to interpret than when this locus was used on Lake Malawi species (Chapter 2 and 3). The results from UNH130 were occasionally difficult to interpret as well as either no or only a single allele amplified. Locus M11 did not always provide interpretable information and has therefore been used in a limited number of broods in this study.

4.4.2 Allele Frequency distributions

Genotype data from adults used in the study and other individuals from the laboratory population reveals that eight alleles are present at UNH130 and nine alleles are present at UME003. Figure 4.3 shows there are six species specific alleles at locus UNH130 and seven species specific alleles at UME003. Both species have a single allele in common at each locus. The frequency distributions for each locus are not continuous as the alleles exhibit a large amount of size variation; UNH130 alleles range in size from 180 – 266 bp and UME003 has a size range of 199 – 257 bp. Alleles present in the *P. nyererei* population are at the extremes of the size range for both loci whereas *P. pundamilia* alleles are more similar in size.

Figure 4.3. Allele frequency distributions for the *P. pundamilia* and *P. nyererei* laboratory populations at a) UNH130 and b) UME003. At UNH130, each species has three unique and one shared alleles. The situation is similar for UME003 where both species also share one allele, but the remainder are species specific.



4.4.3 *Mate Choice of* P. pundamilia *and* P. nyererei *females*

The genotypes of parents and offspring can be found in the Appendix to this chapter. A summary of the results of the paternity assignment may be seen in Table 4.3. Out of a total of eight *P. pundamilia* broods tested (all from different females), seven were sired by the *P. pundamilia* male present at the time. However, allele sharing by the hybrid and *P. pundamilia* male meant that it could not be ascertained whether three of these

broods were the product of a sole mating with the *P. pundamilia* male or multiple paternity. One *P. pundamilia* (B2, Trial 2.1) female mated with a *P. nyererei* male and this occurred under a non-partioned design where behavioural observations indicated the almost total dominance of the *P. nyererei* male in the tank. At the time of spawning of this female, the *P. nyererei* male had extended its territory into almost the entire tank until the *P. pundamilia* male defended an area of only 20cm in diameter (O. Seehausen, pers. com.). This study generated five *P. nyererei* broods from four different females. Of these five, two appeared to be sired by the *P. pundamilia* male present in the tank. Unfortunately only one individual was preserved from one of the broods (R9 Trial 1.2) so it is impossible to know if this brood was of mixed paternity. There are three instances of possible multiple paternity among *P. pundamilia* and *P. nyererei* broods, all from Trial 2.1. In each case the potential for multiple paternity is suggested by a single individual in each brood with an ambiguous genotype. In one case this is due to the presence of a putative null allele (brood B2), and in the two others the ambiguity is caused by conflicting information provided by the two loci (B3 and R2?).

4.4.4 Mate Preference of Hybrid females

Eleven different broods from ten different hybrid females were partially typed for this study (Table 4.4). Of these, three were sired by a *P. nyererei* male, seven were sired by a *P. pundamilia* male and one was sired by both. Uncertainty in paternity assignment in one brood (2H1) was due to an apparent conflict between loci; however this conflict may be due to the presence of a null allele at locus UNH130. The only other individual with uncertain paternity only amplified at locus UNH130; the *P. pundamilia* male and *P. nyererei* male shared an allele at this locus.

4.4.5 BROOD simulations

Mean N* (the number of offspring that should be typed from each brood before all parental alleles are detected) varied depending on both the size of brood and the number of sires and their relative contributions (Table 4.3). It can be seen that a higher number of males with unequal contributions means that more individuals from each brood need to be typed. This is also reflected in the histograms of mean N* (Figure 4.4) that exhibit a more skewed distribution as paternal contributions become less equal.

Table 4.3. N* statistics generated by BROOD for various brood sizes and paternal contributions.

N in brood	Number of sires and relative contributions	Mean N*	SD N*	95% N*
50	3 (1:1:1)	15.19	5.36	25.47
50	2 (2:1)	10.51	4.92	19.93
50	3 (4:1:1)	16.15	7.69	29.72
40	3 (1:1:1)	15.59	5.13	24.7

Figure 4.4. Histograms of cumulative values of N* based on 100 resamplings of 100 simulated broods. Mean N* and the upper 95% CI are marked. A) Results from broods of 50 individuals sired equally by three males. B) Results from broods of 50 individuals sired by 3 males in a 4:1:1 ratio.

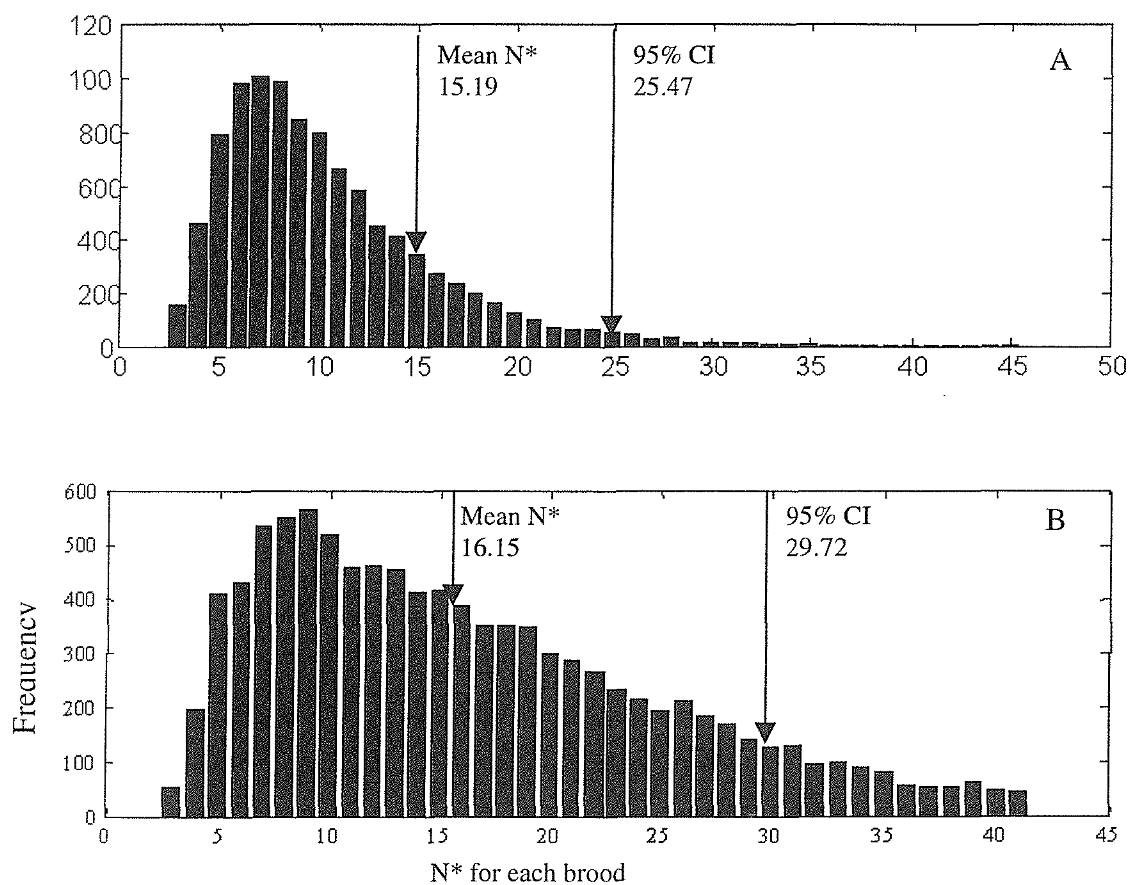


Table 4.4. Summary of results for *P. pundamilia* (B) and *P. nyererei* (R) females from the four trials. Paternity of offspring was categorised as ambiguous if PCR products from one or more loci failed to discriminate between males. The main source of ambiguity resulted from *P. pundamilia* and hybrid males sharing alleles. Another source of ambiguity was the occasional amplification of none or a single allele in offspring either due to errors in PCR or due to the presence of null alleles.

Trial	Female	N in brood	Males Present	Loci used (N typed)			N not typed at any locus	N ambiguous	Paternity			
				UNH130	UME003	M11			B	B OR H	R	?
1.1	B1	34	B1, R1, H1	31	31		3	5	26	5[‡]		
	B2	4	B1, R1, H1	3	3		1	0	3			
1.2	B?1	31	B7s, R4s	28	30		0	0	31			
	R1	19	B7s, R3s	18	18		1	2	14			2 R/B
	R2	25	B7s, R3s	24	24		0	0			25	
	R9	1*	B7s, R3s	1	1		0	0	1[†]			
2.1	B1	43	B1, R1, H2	41	41		0	0	43			
	B2	51	B2, R1, H2	50	50		0	2	0		48[†]	1 R/B
	B3	48	B4, R1, H2	42	31		3	19	26	18[‡]		1 R/B[a]
	B4	65	B4, R1, H2	62	54		1	18	45	18[‡]		
	R?2	51	B4, R1, H2	41	50		0	1			50	1 R/B[a]
2.2	2B	30	B1s, R2s	30	28		0	0	30			
	2R?2	34	B1s, R2s	34		34	0	0			34[†]	

Notes on Table 4.3.

*- only a single individual of this brood was preserved as offspring were lost before they were stripped from the female

[†]- a null allele is suspected in these broods at locus UNH130, originating from the *P. nyererei* population

[‡] - the mother, *P. pundamilia* or hybrid males shared alleles at the two loci used making it impossible to assign paternity with certainty

a – paternity cannot be assigned with certainty as the two loci give conflicting results

Table 4.5. Results of partial typing of broods from 11 F_1 hybrid females from trial 2.1 and 1 hybrid female from Trial 2.2. Ambiguities arose from allele sharing by males or PCR errors and null alleles. The brood from female 2H2 represents one definite case of multiple paternity.

Trial	Female	Males Present	N in brood	N Typed	Loci Used (N typed)			N ambiguous	Paternity			
					UNH130	UME003	M11		B	B or H	R OR H	R
2.1	2H1	B1, R1, H2	14	10	8		9	8			8[†]	2
	2H2	B3, R1, H2	26	17	17		15	0	2			15
	2H3	B3, R1, H2	11	10	10	9		0	10			
	2H4	B4, R1, H2	17	15	15	15	14	0	15			
	2H5	B4, R1, H2	32	19	17	19		0				19
	2H6*	B4, R1, H2	34	33	31	33		1	32			
	2H7	B4, R1, H2	28	15	13	14		0	15			
	2H9	B4, R1, H2	18	16	16	16		0	16			
	2H11	B4, R1, H2	33	14	10		13	0				14
	2H12	B4, R1, H2	28	16	13	15		1	15	1[‡]		
2.2	'2R1'*	B1s, R2s	50	44	44	42		0	44			

Notes on Table 4.4.

*** - '2R1' and 2H6 refer to the same female which was used in both Trial 2.1 and 2.2 and was identified in Trial 2.2 as a *P. nyererei* female. In both trials the entire brood from this female was typed. However, the two locus microsatellite genotype of this female suggested she was misidentified and was actually a hybrid.**

[†]- a null allele is suspected in this brood – if this is the case then all 10 of the individuals typed were sired by the *P. nyererei* male.

[‡]- this individual amplified at only one locus. The hybrid and *P. pundamilia* males shared an allele at this locus therefore the paternity of this individual cannot be assigned with certainty.

4.5 Discussion

4.5.1 Utility of Microsatellite Markers

The loci chosen allowed unambiguous assignment of paternity in most cases. In almost all cases, *P. nyererei* and *P. pundamilia* males could be distinguished with certainty due to the fact that six out of eight (UNH130) and seven out of nine (UME003) alleles were species specific. However, if a hybrid male was present (Trial 1.1 and 2.1) it was often not possible to distinguish between all three males. This was due to the relatively low intraspecific variation of the loci used. As a comparison, a survey of 24 wild caught *M. zebra* individuals revealed 22 alleles at locus UNH130 and 15 alleles at UME003 (data from Chapter 2). The differences in locus variabity between the Victorian and Malawi species probably reflect two facts. Obviously captive populations suffer a loss of genetic variability as they are founded by a small number of individuals. Secondly, both loci were originally isolated from non-Victorian cichlids and one would expect a decrease in variability to be observed in more distantly related species. While the hybrid male cannot be ruled out as contributing to some broods (Table 4.3) the fact that the 'unique' hybrid male allele (i.e. the allele not shared with any other male present) has not been detected in any of the broods in question suggests it was not successful in siring any broods. However, a more variable locus would be required to confirm this. A lack of any PCR products and the occurrence of improbable homozygotes (for example where the mother does not share any alleles with the potential fathers) indicate the presence of null alleles. Such a situation exists at locus UNH130 where a null allele is apparent in several broods (B3 and B4 Trial 2.1 and 2H1, see tables 4.5d and 4.5e in appendix). Null alleles can introduce error into paternity assignment experiments due to false exclusion of potential fathers (Jones *et al*, 1998). Future studies involving microsatellite based paternity assignment in these species should either avoid UNH130 or redesign the primers to enable amplification of all alleles.

The results of this study also may have implications for managing laboratory stock populations to maintain levels of genetic diversity. *P. pundamilia* and *P. nyererei* are not endangered species, but a small founding population (in the case of *P. pundamilia*

<10 individuals and *P. nyererei* ~20 individuals) coupled with the reduced effective population size associated with reproductive success of a few dominant males allows for a rapid loss of diversity. The two *Pundamilia* populations studies here have been captive bred for nine years and approximately six to eight generations. Fiumera *et al* (2000) found that 19% of initial microsatellite alleles were lost within four generations of captive breeding in the endangered species *Prognathochromis perrieri*. The authors recommended the periodic removal of dominant males from the breeding population in order to allow other males to reproduce.

4.5.2 Non Assortative Mating?

The microsatellite typing revealed three broods that were the product of interspecific matings. One mating (female B2, Trial 2.1) occurred in a non-partitioned experimental design when the *P. nyererei* male (the dominant sire of the brood) was observed to be almost completely behaviourally dominant over the *P. pundamilia* male. In situations like this the female may not be able to mate with the male she prefers due to male-male interactions (Turner *et al*, 2001). However, two other broods were the result of matings between *P. nyererei* females and a *P. pundamilia* male, and these occurred under a partitioned design that permitted free choice for the females. In these cases one may conclude that there are no complete pre-zygotic barriers to hybridisation between these two species, or that the experimental design was flawed in some way that distorts female choice. More specifically, an explanation for the heterospecific matings is that the *P. pundamilia* males used were often larger than the *P. nyererei males*; this was the case in 10 out of the 13 pairings. The largest size difference occurred in the B7s/R3s pair used in Trial 1.2 (91mm compared to 84mm) and both heterospecific matings of *P. nyererei* females when these were the only males in the tank. The first explanation of a lack of reproductive isolation seems unlikely as previous mate choice experiments have found assortative mating in situations of free female choice (Seehausen, 1997). In addition, hybrids are not normally observed in the natural environment of these species except in areas of particularly poor visibility (Seehausen *et al*, 1997a). Turner *et al* (2001) state that it is difficult to demonstrate non-assortative mating in laboratory tests of mate choice as the environment under which females are choosing their mates may be very different to the natural conditions the fish have evolved under. One aspect of this is the confined environment of the experimental tank and the limited choice of

males available to females. The results of this experiment may indicate that pre-zygotic isolation is very fragile in this species pair and breaks down when no conspecific male is available which matches the heterospecific male (in terms of size or activity). Whether any of the factors discussed above is the correct explanation for the heterospecific matings should be tested in further experiments using larger tanks and more than one male from each species is available to choose from. The results of mate choice tests such as this are ideally backed up by data on the habitat use and ecology of the species in question, as well as allele frequency data for natural populations at a number of loci in order to estimate the amount of gene flow between them.

4.5.3 Mate Preference of F1Hybrid females.

Seven of the eleven broods produced by the hybrid females were the result of matings with a *P. pundamilia* male. Multiple paternity of broods cannot be ruled out due to the fact that only 10-20 individuals from each brood were typed. This however provides a good indication of mate preference and does detect multiple paternity in one case (brood 2H2). Four broods were the result of hybrid females mating with the *P. nyererei* male. Of these, one was of mixed paternity. Two of the *P. nyererei* sired broods were produced under a partitioned design; the *P. nyererei* male was not behaviourally dominant in these cases or in the two matings that took place under an open experimental design. These results suggest that F1 hybrid female mate preference lies between the preferences of the two parental species. If the parental species are fixed for preference alleles, the hybrid female behaviour is consistent with incomplete dominance of 'blue' (*P. pundamilia*) preference. The results of this study concur well with behavioural indices of hybrid female preference. F1 hybrid females did not exhibit a significant preference for *P. nyererei* or *P. pundamilia* males but were observed to respond slightly more to the *P. pundamilia* males (O. Seehausen, pers. com.). A comparison of the two indices of mate preference (behavioural and microsatellite based) reveals an extremely similar pattern between the two parental species and the F1 hybrids (Figure 4.5).

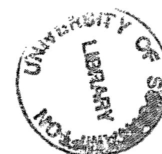
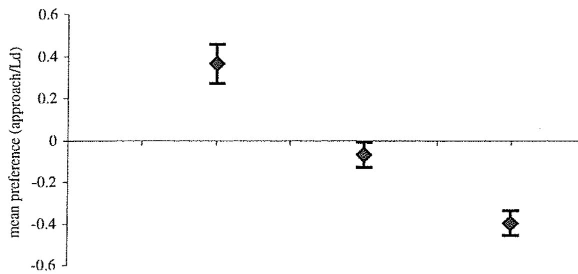
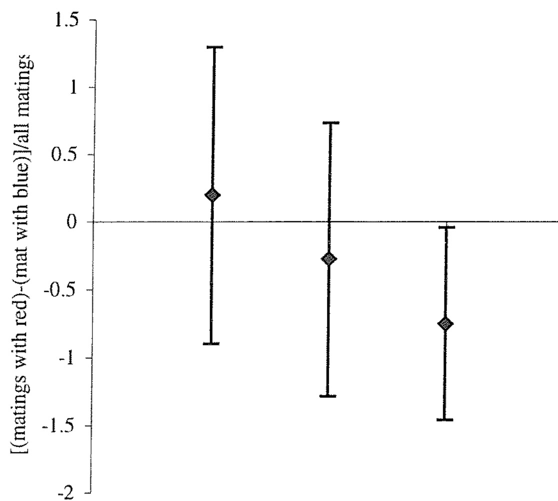


Figure 4.5. Comparison of behavioural and microsatellite based indices of mate preference in *P. pundamilia*, *P. nyererei* and F1 hybrid females. Error bars indicate two standard deviations. Behavioural data were collected by O. Seehausen and R. Fraser.



a) Mate preference of females based on a behavioural parameter [(proportion of red male "lateral display" that elicited female "approach")-(proportion of blue male "lateral display" that elicited female "approach")]. From left: *P. nyererei*, F1 hybrids, *P. pundamilia*.



b) Mate preference of females based on microsatellite determined paternity of offspring. From left: *P. nyererei*, F1 hybrids, *P. pundamilia*.

4.5.4 BROOD Simulations and Sample Size

From Table 4.3, it appears that around 25 members of a brood of 50 offspring should be typed in order to be 95% confident that all parental alleles have been detected, if three males have an equal contribution to the paternity of a brood. When three males provide unequal (4:1:1) contributions to the paternity of a brood the number of offspring requiring typing increases to around 29 (brood size of 50). If only two males sire a brood, then the number of offspring that should be typed decreases to around 19.

Based on these results, broods B1 (Trial 1.1), B3 and B4 (Trial 2.1) are almost certainly the result of a single mating with a *P. pundamilia* male as the 'unique' allele belonging to the hybrid male is not detected, even after the whole brood has been typed. It is difficult to interpret the results of the simulations in terms of the minimum number of offspring that should be typed from each brood, because this requires knowledge of the proportion of offspring sired by each male in the case of multiple paternity. For tests of mate preference, typing up to 29 members (58%) of a brood of 50 individuals should suffice. Of the hybrid broods was typed, with the exception of broods 2H7, 2H11 and 2H12, the proportion typed was equal or greater than 58%, therefore one may assume that the results for the sample of the brood typed reflect the paternity of the whole brood. However, no cases of multiple paternity by three males were detected during all the trials, but there are three cases where a second male may have sired a single individual from a brood. Instances such as these may not be detected unless the whole brood was typed, but in tests of mate preference (rather than absolute mate choice), detecting a single individual with alternate paternity is not important. It may therefore be appropriate to type a small number of individuals from each brood (for example 19 offspring from a brood of 50) as this will still provide the same information with reduced effort.

4.5.5 Conclusions

This study has indicated the utility of microsatellite loci originally isolated from non Victorian cichlids in the study of species of *Pundamilia*, although the results seem to indicate the presence of one or more null alleles at locus UNH130 in the *P. nyererei* population. The importance of experimental design that eliminates male dominance problems is also highlighted, as is the importance of the quality (size and activity) of the males used. Both microsatellite typing of offspring and behavioural indices of mate preference indicate that F1 hybrid females studied have intermediate mate preferences relative to the parental species. This refutes the assumption that single locus complete dominance of female preference traits is necessary for sympatric speciation by sexual selection as proposed in the model of Turner and Burrows (1995).

4.6 REFERENCES

Booton GC, Kaufman L, Chandler M, Ogto-Ohwayo R, Duan W and Fuerst PA (1999) Evolution of the Ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake Victoria Region. *Molecular Phylogenetics and Evolution* 11: 273-282

Dewoody JA, Dewoody YD, Fiumera AC and Avise JC (2000) On the number of reproductives contributing to a half sib progeny array. *Genetic Research* 75 : 95-105

Fiumera AC, Parker PG and Fuerst PA (2000) Effective population size and mainenance of genetic diversity in captive bred populations of a Lake Victoria cichlid. *Conservation Biology* 14: 886-892

Greenwood PH (1984) African cichlid fishes and evolutionary theories. *In Evolution of Fish Species Flocks* (eds Echelle AA and Kornfield I) pp141-154

Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD Ngobi G, Beuning K, Ssemmanda I and McGill JW (1996) Late Pleistocene dessication of Lake Victoria and rapid evolution of cichlid fishes. *Science* 273 1091-1093

Jones AG, Stockwell CA, Walker D and Avise JC (1998) The molecular basis of a microsatellite null allele from the white sands pupfish. *Journal of Heredity* 89 : 339-342

Knight ME, Turner GF, Rico C, van Oppen MJH and Hewitt GM (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology* 7 : 1605-1610

Martens K (1997) Speciation in ancient lakes. *Trends in Ecology and Evolution* 12: 177-182

Meyer A, Kocher TD, Basasibwaki P and Wilson AC (1990) Monophyletic origin of Lake Victoria cichlids suggested by mitochondrial DNA sequences. *Nature* 347 : 550-553

Nagl S, Tichy H, Mayer W, Takeszaki N, Takhhata N and Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proceedings of the Royal Society of London B* 267: 1049-1061

Sage RD, Loiselle PV, Basasibwaki and Wilson AC (1984) Molecular versus morphological change among cichlid fishes of Lake Victoria. In *Evolution of Fish Species Flocks* eds AA Eschelle and I Kornfield. University of Maine Press.

Seehausen O (1996) *Lake Victoria Rock Cichlids.* Verduijn Cichlids, Germany.

Seehausen O and Bouton F (1998) The community of rock dwelling cichlids in Lake Victoria. *Bonn. Zool. Beitr* 47

Seehausen O and van Alphen JJM (1998) The effect of male colouration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioural Ecology and Sociobiology* 42 : 1-8

Seehausen O and van Alphen JJM (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecology Letters* 2: 262-271

Seehausen O, Lippitsch E, Bouton N, Zwennes H (1998b) Mbipi, the rock dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei*). Ichthyological Explorations of Freshwaters* 9 : 129-228

Seehausen O, van Alphen JJM and Lande R (1999) Colour polymorphism and sex ration distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecology Letters* 2: 367-378

Seehausen O, van Alphen JJM, Witte F (1997a) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277: 1808-1811

Seehausen O, Witte F, EF Katunzi, Smits J and Bouton N (1997b) Patterns of the remnant cichlid fauna in southern Lake Victoria. *Conservation Biology* 11: 890-904

Seehausen O, Witte F, van Alphen JJM and Bouton N (1998) Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *Journal of Fish Biology* 53 (SA) 37-55

Turner GF, Seehausen O, Knight ME, Allender CJ and Robinson RL (2001) How many species of cihlid fishes are there in African lakes? Molecular Ecology

van Oppen MJH, Turner GF, Rico C, Robinson RL, Deutsch JC, Genner MJ and Hewitt GM (1998) Assortative mating among rock dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Molecular Ecology* 7: 991-1001

Witte-Maas and Witte F (1985) *Haplochromis nyererei*, a new cichlid fish from Lake Victoria named in honour of Mwalimu Julius Nyerere, President of Tanzania. Brill publishers, Leiden 13pp

4.7 Appendix

Table 4.5 a, b, c, d and e. Genotypes of adults and resulting offspring from the four trials. Only one allele is indicated for homozygotes. Paternity was assigned on the basis of agreement between two loci; if the two loci gave conflicting results then paternity is assigned as ambiguous, however if only one locus of the pair was unclear (e.g. only one allele amplified) and was not in conflict with the other locus then paternity was assigned with certainty. An 'f'' after an allele meant it had a very low intensity (<50 relative fluorescence units) on the ABI gel.

Table 4.5a. Trial 1.1

Individual	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male	**208**	**234**	**231**	**227**	**219**		
R1 male	**180**		**199**	**217**	**214**		
H1 male	**234**		**225**	**239**	**210**	**219**	
B1 female	**224**	**234**	**231**	**239**	**219**		
B1/A	208	234	227	239			B
B1/B	224	234	231				B
B1/C	208	234	231				B
B1/D	224	234	231				B
B1/E	208	234	227	239			B
B1/F	234		231	239			B/H
B1/G	208	234	231				B
B1/H	234		227	239			B
B1/I	234		227	239			B
B1/J	224	234	231	239			B/H
B1/K	208	224	231				B
B1/L	224	234	231				B
B1/M	208	224	231				B
B1/N	234		227	231			B
B1/O	208	224	231	239			B
B1/P	208	234	227	239			B
B1/Q	208	234	231				B
B1/R	224	234	231	239			B/H
B1/S	224	234	231	239			B/H
B1/T	208	234	231	239			B
B1/U	224	234	231				B
B1/V	234		227	239			B
B1/W	234		231				B
B1/X	208	234	231				B
B1/Y	208	234	231	239			B
B1/Z	234		231	239			B/H
B1/AA	208	224	231				B
B1/BB	234		227	239			B
B1/CC	224	234	231				B
B1/DD	224	234	231				B
B1/EE	234		231				B
B1/FF							?
B1/GG							?
B2	**224**	**234**	**227**	**239**	**219**		
B2/1	224	234	231	239			B
B2/2	208	224	227	239			B
B2/3	208	224	231				B
B2/4							

Table 4.5b. Trial 1.2. The brood of female R1 in this trial proved difficult to amplify and results were therefore only obtained for one locus. In the absence of any other data, paternity has been tentatively assigned using only locus UNH130. All except one individual of the brood of female R9 were lost before DNA extraction.

Individual	Males Present	UNH130	UNH130	UME003	UME003	Paternity
B7s male		**208**	**232**	**223**	**231**	
R3s male		**180**		**199**	**217**	
R4s Male		**180**	**266**	**217**		
B?1	B7s, R4s	**234**		**231**	**239**	
B?1/1		208	234f	231	239	B
B?1/2		208	234	231	239	B
B?1/3		208	234	231		B
B?1/4		232	234	231	223	B
B?1/5		208	234	227	239	B
B?1/6		232	234	223	239	B
B?1/7		232	234	231	239	B
B?1/8		208	234	231	223	B
B?1/9		232	234	223	239	B
B?1/10		208	234	231		B
B?1/11		208	234	223	239	B
B?1/12		232	234	231		B
B?1/13		232	234	231		B
B?1/14				231	239	B
B?1/15		232	234	231	239	B
B?1/16		208	234	231	239	B
B?1/17		232	234	231	239	B
B?1/18		208	234	231	239	B
B?1/19		208	234	223	239	B
B?1/20		208	234f			B
B?1/21		208	234	223	239	B
B?1/22		232	234	231	239	B
B?1/23		208	234	231		B
B?1/24				231		B
B?1/25		208	234	231	223	B
B?1/26		208	234	231		B
B?1/27				231		B
B?1/28		232	234	231	223	B
B?1/29		232	234	231		B
B?1/30		208	234	231		B
B?1/31		232	234	231	239	B
R1	B7s, R3s	**180**		**217**	**257**	
R1/1		180	208	223	257	B
R1/2		208	180	231	257	B
R1/3		180	232f	231		B
R1/4		208	180	231	257	B
R1/5		180f	232	223	257	B
R1/6		180.4f		231	257	B/R
R1/7		180f	208	223	257	B
R1/8		180		217	231	B/R

Individual	Males Present	UNH130	UNH130	UME003	UME003	Paternity
B7s male		**208**	**232**	**223**	**231**	
R3s male		**180**		**199**	**217**	
R4s Male		**180**	**266**	**217**		
R1/9		180	232	217	231	B
R1/10		232		217	223	B
R1/11		208		217	223	B
R1/12						?
R1/13		180	232			B
R1/14		208	180	223	257	B
R1/15		180	232	223	257	B
R1/16		180	208	217	223	B
R1/17		180	232	217	231	B
R1/18		180f	232			B
R1/20		180	232	217	231	B
R2	B7s, R3s	**180**	**266**	**217**	**257**	
R2/1		180	266	217		R
R2/2		180		217		R
R2/3		180		217	257	R
R2/4		180	266	199	217	R
R2/5		180		199	257	R
R2/6		180		199	217	R
R2/7				199	217	R
R2/8		180	266	199	217	R
R2/9		180				R
R2/10		180	266	217		R
R2/11		180	266	199		R
R2/12		180		217		R
R2/13		180		217	257	R
R2/14		180		217		R
R2/15		180		199	257	R
R2/16		266		199	257	R
R2/17		180		217	257	R
R2/18		180	266	199	257	R
R2/19		266		199	217	R
R2/20		180		217		R
R2/21		266		199		R
R2/22		180		217		R
R2/23		180	266	199	217	R
R2/24		180		199	217	R
R2/25		180	266	199	217	R
R9		**180**		**217**	**257**	
R9/1		208		223	257	

Table 4.5c. Trial 2.1.

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		

B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2B1	**B1, R1, H2**	**208**	**224**	**231**		
2B1/1		224		231		B
2B1/2		208	224			B
2B1/3		208	224	231		B
2B1/4		208	224	231		B
2B1/5				231		B
2B1/6		208	224	231		B
2B1/7		208	224	231		B
2B1/8		208	224	231		B
2B1/9		208		231		B
2B1/10		208	224	231		B
2B1/11		208		231		B
2B1/12		208	224	231		B
2B1/13		*211*	224	231		B
2B1/14		208		231		B
2B1/15		208		231		B
2B1/16		208		231		B
2B1/17		208	224	231		B
2B1/18		208	224	231		B
2B1/19		208	224	231		B
2B1/20		208	224	231		B
2B1/21		208	224	231		B
2B1/22		208		231		B
2B1/23		208	224	231		B
2B1/24		208		231		B
2B1/25		208		231		B
2B1/26		208	224	231		B
2B1/27		208	224	231		B
2B1/28		208	224	231		B
2B1/29		208	224	231		B
2B1/30		208		231		B
2B1/31		208		231		B
2B1/32		208	224	231		B
2B1/33		208	224	231		B
2B1/34		224		231		B
2B1/35		224		231		B
2B1/36		208	224	231		B
2B1/37		224		231		B
2B1/38		208	224	231		B
2B1/39		224		231		B
2B1/40		208	224	231		B
2B1/41		224		231		B
2B1/42		208		231		B
2B1/43				231		B
2B2	**B2, R1, H2**	**224**	**234**	**231**	**239**	
2B2/1		224		217	231	R?
2B2/2		234		199	239	R?
2B2/3		180	224	199	231	R

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		
B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2B2/4		234		217	239	R?
2B2/5		180	234	199	231	R
2B2/6		224		199	239	R?
2B2/7		180	234	199	239	R
2B2/8		180	234	199	239	R
2B2/9		180	234	199	239	R
2B2/10		180	234	199	231	R
2B2/11		224		199	239	R?
2B2/12		224		199	239	R?
2B2/13		180		217	231	R
2B2/14		224		217	231	R?
2B2/15		180	234	199	231	R
2B2/16		180	224	217	239	R
2B2/17		180	224	217	231	R
2B2/18		224		199	239	R?
2B2/19		224		217	239	R?
2B2/20		224		199	231	R?
2B2/21		224		231		B/R
2B2/22		224		217	231	R?
2B2/23		180	224	217	231	R
2B2/24		180	234	217	231	R
2B2/25		234		217	239	R?
2B2/26		180	224	217	231	R
2B2/27		180	224	199	231	R
2B2/28		180	234	199	231	R
2B2/29		180	234	231		R
2B2/30		180	234	217	239	R
2B2/31		180	234	231		R
2B2/32		180	234	217	239	R
2B2/33		234		217	239	R?
2B2/34		234		199	239	R?
2B2/35		224		199	239	R?
2B2/36		234		217	231	R?
2B2/37		180	224	199	239	R
2B2/38		180	224	217	239	R
2B2/39		234		199		?
2B2/40		180	224	217	239	R
2B2/41		224		217	231	R?
2B2/42		224		217	239	R?
2B2/43		224		217	231	R?
2B2/44		180	234	199		R
2B2/45		224		217	239	R?
2B2/46		224		217	231	R?
2B2/47		180	234	217	231	R
2B2/48		180f	234	199	231	R
2B2/49		180	234	199	231	R

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		
B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2B2/50		180	234	217	231	R
2B2/48B				199	231	R
2B3	**B4, R1, H2**	**208**	**234**	**231**	**239**	
2B3/1						?
2B3/2		208	234			B or H
2B3/3		208	234	231		B
2B3/4		208	234	231		B
2B3/5		208	234			B OR H
2B3/6				231		B
2B3/7		234		231		B
2B3/8		208	234	231		B
2B3/9						?
2B3/10		208				B
2B3/11		208	234	231	239	B OR H
2B3/12		234				B OR H
2B3/13		208		231		B
2B3/14		234		231		B
2B3/15		234				B OR H
2B3/16		224	234	231	239	B OR H
2B3/17		208	234	231	239	B OR H
2B3/18		234		231		B
2B3/19		208	234	231		B
2B3/20				231		B
2B3/21		208	234	231		B
2B3/22		208		231		B
2B3/23		234		199	231	B/R
2B3/24		208		231		B
2B3/25		208	234	231	239	B OR H
2B3/26		208		231	239	B
2B3/27		208	234	231		B
2B3/28		234		231		B
2B3/29		234		231	239	B OR H
2B3/30		208	234			B OR H
2B3/31		208	234	231	239	B OR H
2B3/32		208	234	231	239	B OR H
2B3/33		208	234	231		B
2B3/34		208	234	231	239	B OR H
2B3/35		208	234	231	239	B OR H
2B3/36		208	234	231	239	B OR H
2B3/37				231		B
2B3/38		208	234	231		B
2B3/39		208	234	231		B
2B3/40		208	234	231		B
2B3/41		234		231		B
2B3/42		208		228		B
2B3/43						?

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		
B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2B3/44		208		231		B
2B3/45		208	234	231	239	B OR H
2B3/46		208		231	239	B
2B3/47		208	234	231	239	B OR H
2B3/48		208	234	231	239	B OR H
2B4 (2H8)	**B4, R1, H2**	**224**	**234**	**231**	**239**	
2B4/1		208	234	231	239	B OR H
2B4/2		234		231	239	B OR H
2B4/3		208	224	231	239	B
2B4/4		208	224	231		B
2B4/5		208	224	231	239	B
2B4/6		208	234	231		B
2B4/7		234				B OR H
2B4/8		208	224			B
2B4/9		208	234	231		B
2B4/10		208	224	231		B
2B4/11		234		231		B
2B4/12		224	234	231		B
2B4/13		224	234	231		B
2B4/14		208	224	231	239	B
2B4/15		208	224	231		B
2B4/16		208	224	231	239	B
2B4/17		208	224	231	239	B
2B4/18		224	234	231		B
2B4/19		224	234	231	239	B OR H
2B4/20		234		231		B
2B4/21				231	239	B OR H
2B4/22		234		231		B
2B4/23		234		231	239	B OR H
2B4/24		208	224	231	239	B
2B4/25		208	234	231	239	B
2B4/26		234		231		B
2B4/27		224	234	231		B
2B4/28		208	224	231		B
2B4/29		208	224			B
2B4/30		208	234	231		B
2B4/31		208	234	231		B
2B4/32		224	234	231	239	B OR H
2B4/33		224	234	231	239	B OR H
2B4/34		208	234	231		B
2B4/35		234		231		B
2B4/36		234		231		B
2B4/37		208	224	231		B
2B4/38		208	224	231	239	B
2B4/39		234		231	239	B OR H

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		
B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2B4/40		234		231	239	B OR H
2B4/41		224	234	231	239	B OR H
2B4/42		224	234			B OR H
2B4/43		234		231		B
2B4/44		208	234	231		B
2B4/45		208	224	231		B
2B4/46		208	224	231	239	B
2B4/47		224	234	231		B
2B4/48		224	234	231		B
2B4/49		234		231	239	B OR H
2B4/50		224	234	231	239	B OR H
2B4/51				231	239	B OR H
2B4/52		208	234	231	239	B
2B4/53		224	234	231		B
2B4/54		234		231		B
2B4/55		234		231		B
2B4/56		208	234	231	239	B
2B4/57		208	224	231	239	B
2B4/58		234				B OR H
2B4/59		208	234			B
2B4/60		208	224			B
2B4/61		224	234			B OR H
2B4/62		234				B
2B4/63		208				B
2B4/64						?
2B4/65		224	234			B OR H
2R?2	**B4, R1, H2**	**180**		**199**	**217**	
2R?2/1		180		217	231	B/R
2R?2/2		180		199		R
2R?2/3		180		199		R
2R?2/4		180				R
2R?2/5		180		217		R
2R?2/6		180		199		R
2R?2/7		180		199		R
2R?2/8		180		217	199	R
2R?2/9		180		199		R
2R?2/10		180		199	217	R
2R?2/11		180		217		R
2R?2/12		180		199		R
2R?2/13		180		199		R
2R?2/14		180		217		R
2R?2/15		180		199	217	R
2R?2/16		180		199		R
2R?2/17				199	217	R
2R?2/18		180		199	217	R

Individual	Males present	UNH130	UNH130	UME003	UME003	Paternity
B1 male		**208**	**234**	**231**		
B2 male		**208**	**224**	**231**		
B3 Male		**224**	**232**	**223**	**231**	
B4 male		**208**	**234**	**231**		
H2 male		**234**	**266**	**225**	**239**	
R1 male		**180**		**199**	**217**	
2R?2/19		180		199		R
2R?2/20		180		200		R
2R?2/21		180		217		R
2R?2/22		180		199	217	R
2R?2/23		180		199		R
2R?2/24		180		199	217	R
2R?2/25				217		R
2R?2/26				217		R
2R?2/27		180		217		R
2R?2/28				199	217	R
2R?2/29				217		R
2R?2/30				199		R
2R?2/31		180		217		R
2R?2/32		180		199		R
2R?2/33		180		199		R
2R?2/34		180		199	217	R
2R?2/35				199		R
2R?2/36		180		199		R
2R?2/37		180		199	217	R
2R?2/38		180		199	217	R
2R?2/39		180		199	217	R
2R?2/40		180		199	217	R
2R?2/41		180		199		R
2R?2/42		180		217		R
2R?2/43		180		199	217	R
2R?2/44				199		R
2R?2/45				199		R
2R?2/46				199		R
2R?2/47		180		199		R
2R?2/48		180		199	217	R
2R?2/49		180		199	217	R
2R?2/50		180		199		R
2R?2/51		180		199	217	R

Table 4.5d. Trial 2.2

Individual	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1s male	**208**	**224**	**231**		**217**	**219**	
R2s Male	**266**		**199**	**225**	**214**		
2B*	**208**	**234**	**199**	**231**			
2B/1	208		231				B
2B/2	208	234	231				B
2B/3	208	224	231				B
2B/4	208		231				B
2B/5	208	224	231				B
2B/6	208		231				B

Individual	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1s male	**208**	**224**	**231**		**217**	**219**	
R2s Male	**266**		**199**	**225**	**214**		
2B/7	208	234	231				B
2B/8	224	234	231				B
2B/9	208	224	231				B
2B/10	208	234	231				B
2B/11	224	234	231				B
2B/12	208	234	231				B
2B/13	208	234	231				B
2B/14	208	224	231				B
2B/15	224	234	231				B
2B/16	208	224	231				B
2B/17	224	234	231				B
2B/18	208	224	231				B
2B/19	208		231				B
2B/20	208	224	231				B
2B/21	208	234					B
2B/22	224	234	231				B
2B/23	208	234	231				B
2B/24	208	224	231				B
2B/25	208	224	231				B
2B/26	224	234	231				B
2B/27	224	234	231				B
2B/28	208	234	231				B
2B/29	208	234	231				B
2B/30	208	224					B
2R?2	**180**	**266f**	**199**	**217**	214		
2R?2/1	180	266			214		R
2R?2/2	180				214		R
2R?2/3					214		R
2R?2/4					214		R
2R?2/5					214		R
2R?2/6	180				214		R
2R?2/7	180				214		R
2R?2/8	266				214		R
2R?2/9	266				214		R
2R?2/10	180	266			214		R
2R?2/11	180				214		R
2R?2/12					214		R
2R?2/13	180	266			214		R
2R?2/14	180				214		R
2R?2/15	180				214		R
2R?2/16	180				214		R
2R?2/17	180	266			214		R
2R?2/18	180				214		R
2R?2/19	180	266			214		R
2R?2/20	266				214		R
2R?2/21	180				214		R
2R?2/22	266				214		R
2R?2/23	180				214		R
2R?2/24	180	266					R

Individual	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1s male	**208**	**224**	**231**		**217**	**219**	
R2s Male	**266**		**199**	**225**	**214**		
2R?2/25					214		R
2R?2/26					214		R
2R?2/27	180				214		R
2R?2/28	180				214		R
2R?2/29	180				214		R
2R?2/30	180	266			214		R
2R?2/31					214		R
2R?2/32	266				214		R
2R?2/33					214		R
2R?2/34	180	266			214		R

Table 4.5e. Paternity of broods from F1 hybrid females, Trial 2.1

Individual	Males Present	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male		**208**	**234**	**231**	**231**	**219**	**219**	
B2 male		**208**	**224**	**231**	**231**	**217**	**219**	
B3 Male		**224**	**232**	**223**	**231**	**210**	**210**	
B4 male		**208**	**234**	**231**	**231**	**219**	**219**	
H2 male		**234**	**266**	**225**	**239**	**210**	**214**	
R1 male		**180**		**199**	**217**	**214**	**214**	
2H1	**B1, R1, H2**	**208**	**208**			**210**	**219**	
2H1/1		175				214	219	R OR H
2H1/2		180				214	219	R OR H
2H1/3		180				210	214	R OR H
2H1/4		208				214	219	R OR H
2H1/5						210	214	R OR H
2H1/6		180	208			214	219	R OR H
2H1/7						210	214	R
2H1/8		208				210	214	R OR H
2H1/9		180				210	214	R OR H
2H1/10		180	208					R
2H2	**B3, R1, H2**	**180**	**224**	**199**	**217**	**210**	**219**	
2H2/1		180		217	217	214	219	R
2H2/2		180	224	217	217	210	214	R
2H2/3		224		199	217	210	214	R
2H2/4		224		217	217	214	219	R
2H2/5		224		217	217	210	210	B
2H2/6		180		199	199	214	219	R
2H2/7		180		199	217	210	214	R
2H2/8		180		199	217			R
2H2/9		180		217	217			R
2H2/10		224		199	199	214	219	R
2H2/11		224		199	217	214	219	R
2H2/12		180	232	217	231	210	219	B
2H2/13		224		199	199	214	219	R
2H2/14		180	180	199	199	210	214	R
2H2/15		224	232	199	223	210	219	B
2H2/16		180	224	199	217	214	219	R

Individual	Males Present	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male		**208**	**234**	**231**	**231**	**219**	**219**	
B2 male		**208**	**224**	**231**	**231**	**217**	**219**	
B3 Male		**224**	**232**	**223**	**231**	**210**	**210**	
B4 male		**208**	**234**	**231**	**231**	**219**	**219**	
H2 male		**234**	**266**	**225**	**239**	**210**	**214**	
R1 male		**180**		**199**	**217**	**214**	**214**	
2H2/17		180	224	199	217	214	219	R
2H3	**B3, R1, H2**	**180**	**208**	**217**	**225**			
2H3/1		208	232	225	231			B
2H3/2		180	224	223	225			B
2H3/3		180	232	223	225			B
2H3/4								?
2H3/5		180	224	225	231			B
2H3/6		208	232	217	231			B
2H3/7		180	224	225	231			B
2H3/8		180	224	223	225			B
2H3/9		180	224					B
2H3/10		180	224	217	231			B
2H3/11		180	224	225	231			B
2H3/12								?
2H4	**B4, R1, H2**	**208**	**208**	**217**	**225**	**210**	**219**	
2H4/1		208		231	225	210	219	B
2H4/2		208		231	225	210	210	B
2H4/3		234	234	231	225	210	210	B
2H4/4								?
2H4/5		234	234	217	231	219	219	B
2H4/6		208		217	231	210	219	B
2H4/7		208	234	217	231	210	219	B
2H4/8		208		231	225	210	210	B
2H4/9		208		231	225	219	219	B
2H4/10		208		231	225	219	219	B
2H4/11		208		231	225	219	219	B
2H4/12		208	234	225	231	219	219	B
2H4/13		208	234	231	225	210	210	B
2H4/14		208	234	231	225	219	219	B
2H4/15		234	234	231	225	210	210	B
2H4/16								?
2H4/17		208	234	217	231			B
2H5		**180**	**208**	**217**	**225**			
2H5/1		180	180	217	225			R
2H5/2		180	180	199	217			R
2H5/3		208		199	225			R
2H5/4		180	180	199	217			R
2H5/5		180	180	217	225			R
2H5/6		180	180	199	217			R
2H5/7		180	208	217	217			R
2H5/8		180	208	217	225			R
2H5/9				199	217			R
2H5/10								?
2H5/11		180	208	199	225			R
2H5/12		180	208	199	225			R
2H5/13		180	180	217	225			R

Individual	Males Present	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male		**208**	**234**	**231**	**231**	**219**	**219**	
B2 male		**208**	**224**	**231**	**231**	**217**	**219**	
B3 Male		**224**	**232**	**223**	**231**	**210**	**210**	
B4 male		**208**	**234**	**231**	**231**	**219**	**219**	
H2 male		**234**	**266**	**225**	**239**	**210**	**214**	
R1 male		**180**		**199**	**217**	**214**	**214**	
2H5/14		180	180	199	225			R
2H5/15		180	208	199	217			R
2H5/16		180	180	199	217			R
2H5/17		208		199	217			R
2H5/18		180	208	199	217			R
2H5/19				199	217			R
2H5/20		208		199	225			R
2H6	**B4, R1, H2**	**186**	**208**	**199**	**217**			
2H6/1		208		199	217			?
2H6/2		186	208	217	231			B
2H6/3		186	208	199	231			B
2H6/4		186	234	199	231			B
2H6/5		208	234	217	231			B
2H6/6		186	208	199	231			B
2H6/7		186	234	199	231			B
2H6/8		186	234	217	231			B
2H6/9		186	208	217	231			B
2H6/10		186	208	199	231			B
2H6/11		186	208	199	231			B
2H6/12		186	234	199	231			B
2H6/13		208		199	231			B
2H6/14		208	234	199	231			B
2H6/15		208		199	231			B
2H6/16		186	208	199	231			B
2H6/17		185	234	199	231			B
2H6/18		185	208	199	231			B
2H6/19		208		217	231			B
2H6/20		185	234	217	231			B
2H6/21		185	234	199	231			B
2H6/22		185	208	217	231			B
2H6/23								?
2H6/24		185	234	217	231			B
2H6/25		208	234	217	231			B
2H6/26		185	208	199	231			B
2H6/27		208		217	231			B
2H6/28				231				?
2H6/29		185	234	199	231			B
2H6/30		185	208	199	231			B
2H6/31		208	234	199	231			B
2H6/32		185	208	199	231			B
2H6/33		208	234	199	231			B
2H6/34				199	231			B
2H7	**B4, R1, H2**	**180**	**224**	**217**	**225**			
2H7/1		224	234	217	231			B
2H7/2		180	208	217	231			B
2H7/3		224	208					B

Individual	Males Present	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male		**208**	**234**	**231**	**231**	**219**	**219**	
B2 male		**208**	**224**	**231**	**231**	**217**	**219**	
B3 Male		**224**	**232**	**223**	**231**	**210**	**210**	
B4 male		**208**	**234**	**231**	**231**	**219**	**219**	
H2 male		**234**	**266**	**225**	**239**	**210**	**214**	
R1 male		**180**		**199**	**217**	**214**	**214**	
2H7/4				225	231			B
2H7/5				217	231			B
2H7/6		234	234	231	231			B
2H7/7		208	224	217	231			B
2H7/8		180	234	217	231			B
2H7/9		208	224	225	231			B
2H7/10		208	224	217	231			B
2H7/11		180	234	217	231			B
2H7/12		208	224	225	231			B
2H7/13		224	234	216	231			B
2H7/14								?
2H7/15		208	224	217	231			B
2H7/16		180	208	231	225			B
2H9	**B4, R1,H2**	**180**	**208**	**217**	**225**			
2H9/1		180	234	216	231			B
2H9/2		180	208	216	231			B
2H9/3		208	234	216	231			B
2H9/4		180	208	225	231			B
2H9/5		180	234	225	231			B
2H9/6		208	234	225	231			B
2H9/7		180	234	216	231			B
2H9/8		180	208	225	231			B
2H9/9		180	234	217	231			B
2H9/10		208	234	217	231			B
2H9/11		208	234	225	231			B
2H9/12		180	234	225	231			B
2H9/13		180	234	225	231			B
2H9/14		180	208	217	231			B
2H9/15		208	208	217	231			B
2H9/16		208	234	225	231			B
2H11		**180**	**180**			**214**	**219**	
2H11/1		180				214		R
2H11/2		180						R
2H11/3		180				214		R
2H11/4						214	219	R
2H11/5						214		R
2H11/6		180				214	219	R
2H11/7						214	219	R
2H11/8		180				214		R
2H11/9		180				214	219	R
2H11/10		180				214		R
2H11/11		180				214		R
2H11/12		180				214		R
2H11/13		180				214		R
2H11/14								?
2H11/15								?

Individual	Males Present	UNH130	UNH130	UME003	UME003	M11	M11	Paternity
B1 male		**208**	**234**	**231**	**231**	**219**	**219**	
B2 male		**208**	**224**	**231**	**231**	**217**	**219**	
B3 Male		**224**	**232**	**223**	**231**	**210**	**210**	
B4 male		**208**	**234**	**231**	**231**	**219**	**219**	
H2 male		**234**	**266**	**225**	**239**	**210**	**214**	
R1 male		**180**		**199**	**217**	**214**	**214**	
2H11/16						214		R
2H12		**180**	**208**	**199**	**217**			
2H12/1		208	234	217	231			B
2H12/2				199	231			B
2H12/3		180	208	217	231			B
2H12/4		180	234	217	231			B
2H12/5		208	208	217	231			B
2H12/6		180	208	199	231			B
2H12/7		180	208	199	231			B
2H12/8		208	234	217	231			B
2H12/9		208		199	231			B
2H12/10		208		217	231			B
2H12/11				217	231			B
2H12/12		180	234					B OR H
2H12/13		208	234	199	231			B
2H12/14		180	208	217	231			B
2H12/15		208	234	199	231			B
2H12/16		180	234	217	231			B

Table 4.5f. Paternity of brood from female 2R1 Trial 2.2, the same female revealed by genotype to be 2H6 in Trial 2.1

Individual	UNH130	UNH130	UME003	UME003	Paternity
B1s male	**208**	**224**	**231**		
R2s Male	**266**		**199**	**225**	
2R1	**186**	**208**	**199**	**217**	
2R1/1	208	224			B
2R1/2	186	224			B
2R1/3	186	208	217	231	B
2R1/4	208		199	231	B
2R1/5	208		199	231	B
2R1/6	208		217	231	B
2R1/7	208		217	231	B
2R1/8	211		217	231	B
2R1/9	208		217	231	B
2R1/10	208	224	217	231	B
2R1/11	208		217	231	B
2R1/12	208	224	217	231	B
2R1/13	208	224	217	231	B
2R1/14	186	208	217	231	B
2R1/15	186	224	199	231	B
2R1/16	186	208	217	231	B
2R1/17	208		199	231	B
2R1/18	208	224	217	231	B

Individual	UNH130	UNH130	UME003	UME003	Paternity
B1s male	**208**	**224**	**231**		
R2s Male	**266**		**199**	**225**	
2R1/19	186	208	199	231	B
2R1/20	208	224	199	231	B
2R1/21	208	224	217	231	B
2R1/23	208		199	231	B
2R1/24	208	224	217	231	B
2R1/25	186	208	217	231	B
2R1/26	186	224	217	231	B
2R1/27	186	224	217	231	B
2R1/28	208	224	217	231	B
2R1/29	186	224	217	231	B
2R1/30	208	224	199	231	B
2R1/31	186	208	199	231	B
2R1/32	208	224	217	231	B
2R1/33	186	208	199	231	B
2R1/34	186	208	199	231	B
2R1/35	208		199	231	B
2R1/36	208		199	231	B
2R1/37	208	224	217	231	B
2R1/38	208		199	231	B
2R1/39	186	208	199	231	B
2R1/40	208	224	199	231	B
2R1/41	186	208	217	231	B
2R1/42	186	224	217	231	B
2R1/43	186	224	199	231	B
2R1/44	186	208	217	231	B

5 Using Microsatellites to Reconstruct the Phylogeny of Lake Malawi Cichlid Fish

5.1 Introduction

As discussed in the General Introduction to this thesis, inter- and intra-lake phylogenies have been reconstructed using a variety of molecular markers. While these have been successful in resolving deep divergences, the relationships within and between clades within lakes have proved more difficult to determine. However, some progress has been made in this area. Nishida (1991) described the polyphyletic origin of the endemic cichlids of Lake Tanganyika using data from allozyme electrophoresis. In contrast, Sage *et al* (1994) were unable to resolve relationships between haplochromine cichlids from Lake Victoria due to a lack of variability in the nine allozyme markers used. These two studies reflect the fact that slow evolving markers such as allozymes are best used to study ancient divergences; the Tanganyikan flock is thought to be at least twenty times older than the Victorian flock (see table 1.1, Chapter 1 for references).

Fast mutating mitochondrial genes have been employed in several phylogenetic studies of African cichlids. The earliest of these studies (Meyer *et al*, 1990; Meyer 1993 – see Figure 1.3, Chapter 1) attempted to determine phylogenetic relationships between the three species flocks. Mitochondrial genes were also sequenced by Idid *et al* (unpublished, some data in Shaw *et al*, 2000) to investigate the intra-lake structure of the Malawi species flock (Figure 5.1). Six clades were found, each with high levels of statistical support. Five of these clades correspond with the five clades determined by Meyer (1993). However, as Figure 5.1 shows, there is very little intra-clade phylogenetic structure. A similar study was carried out on the Lake Victoria super flock by Nagel *et al* (2000) who determined that the species present in modern Lake Victoria do not represent a monophyletic group. Mitochondrial DNA was also used in an attempt to ascertain the phylogenetic relationships between several mbuna species (Moran and Kornfield, 1995). However, the rapid evolution of this group has left shared ancestral polymorphisms caused by incomplete lineage sorting, and this confounds any attempt at phylogenetic analysis.

The extremely rapid divergence that has occurred within the mbuna clade means that markers with high mutation rates are required. Kornfield and Parker (1997) tested a two locus microsatellite data set for the strength of phylogenetic signal it contained. Phenograms were constructed based on matrices of genetic distance between the species used. With just two loci they were able to demonstrate that congeneric species in the mbuna clade clustered together to the exclusion of species from different genera. The results were not perfect in that the relationships between two species of the genus *Melanochromis* were never fully resolved. However, as the authors pointed out, this was not surprising because two loci is an extremely small number with which to reconstruct phylogenetic relationships.

There are several examples of robust phylogenies based entirely on microsatellite data in the published literature. It appears that microsatellites are able to provide information on evolutionary relationships between closely related species, and most of the phylogenies published are of such groupings. Petren *et al* (1999) used data from sixteen loci to estimate the phylogeny of thirteen species of Darwin's finches from the Galapagos Islands. The phylogeny provided useful information about the biogeography and order of divergence within this group of species. Bowcock *et al* (1994) used thirty microsatellite loci to examine relationships between fourteen indigenous populations of humans from five continents. The phylogeny of the *Drosophila melanogaster* species complex was estimated using thirty-nine dinucleotide loci by Harr *et al* (1998).

5.1.1 Aims

Data from up to 39 loci have been used for published microsatellite based phylogenetic studies. However, Kornfield and Parker (1997) demonstrated the phylogenetic potential of a two locus data set on Malawi cichlids. This study aims to explore the phylogenetic potential of data combined from three pre-existing six locus data sets. The biological usefulness of the resulting trees will be assessed using known information on the species included, for example relationships between species determined using other markers (e.g. Figure 5.1) and current taxonomic classifications. Three different genetic distance statistics as well as a variety of tree building methods will be tested.

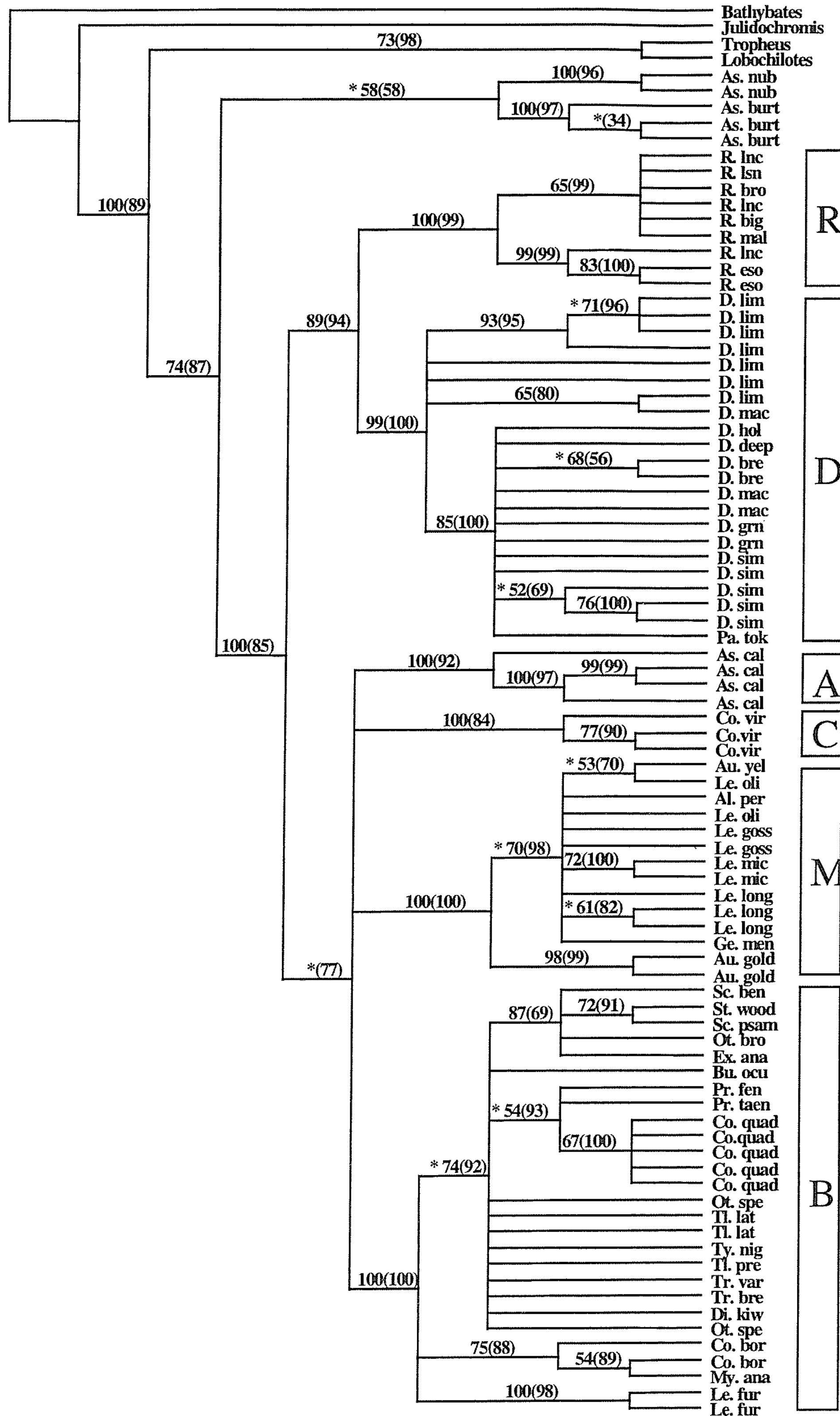


Figure 5.1. Phylogeny of Malawi cichlids based on a combination of NAD2 and D loop sequences. Taken from Idid *et al* (unpublished manuscript). The six clades of Malawi cichlids are marked. R – *Rhamphochromis*; D – *Diplotaxodon*; A – *Astatotilapia*; C – *Copidochromis*; M – mbuna; B – Benthic.

5.2 Methods

5.2.1 Populations Included

Three previously collected datasets were combined for the purpose of phylogenetic analysis. One contained data for mbuna (rocky shore fish) populations located around Nkhata Bay, the second related to mbuna populations at Cape Manulo, Mara Rocks and Ruarwe. The third dataset included three populations of pelagic *Diplotaxodon*, two of *D. limnothrissa* and one of *D. macrops*. Species, population location and sample sizes are given in Table 5.1.The locations of the populations sampled can be seen in Figure 5.2. All three data sets contained allele frequencies at six loci; Pzeb1, Pzeb2, Pzeb3, Pzeb4, Pzeb5 and UNH002. All mbuna samples were typed and analysed on the same ALF (Pharmacia) sequencer at the University of East Anglia. The *Diplotaxodon* samples were typed on the same model of sequencer at the University of Hull. However, to ensure compatibility between the data sets, certain mbuna samples were typed on both machines and the allele sizes compared. For some of the loci it was necessary to adjust the allele size given by the sequencer at Hull in order to ensure compatibility with the data collected at UEA (genotyping and corrections carried out by P. Shaw).

Table 5.1. Identity, collection locality and sample size for the 36 populations included in the phylogenetic analysis. Previously published references pertinent to each data set are as follows. O – van Oppen *et al* (1997), B - Bouteillon (1998), S – Shaw *et al* (2000)

Species	Population Location	Abbreviation	Reference	N
Tropheops species 'Band'	Nkukuti Point	Bandnk	O	53
	Ruarwe	Bandrw	B	24
Tropheops species 'Black'	Nkukuti Point	BLnk	O	77
Tropheops species 'Rust'	Nkukuti Point	Runk	O	51
Tropheops species 'Deep'	Nkukuti Point	Denk	O	61
Diplotaxodon macrops	'39' - South East Arm	Dmac	S	66
D. limnothrissa	'38' – South East Arm	Dlim	S	57
	'4' – Northern offshore	Dl	S	59
Tropheops species 'deep'	Nkukuti Point	Denk	O	61
Metriaclima zebra species	Ruarwe	Gorw	B	40
'gold'	Nkukuti Point	Gonk	O	50
M. callainos	Cape Manulo	Cobcm	B	65
	Chindazowa Point	Cobchi	O	67
	Chirundu Point	Cobchr	O	81
	Mara Rocks	Cobmr	B	62

Species	Population Location	Abbreviation	Reference	N
	Mwafufu	Cobmwa	O	45
	Nkukuti Point	Cobnk	O	121
	Ruarwe	Cobrw	B	57
M. zebra	Cape Manulo	BBcm	B	53
	Chindazowa Point	Bbchi	O	59
	Chirundu Point	BBchr	O	54
	Mara Rocks	BBmr	B	80
	Nkukuti Point	BBnk	O	104
	Ruarwe	BBrw	B	59
Tropheops species 'mauve'	Cape Manulo	Macm	B	37
	Chindazowa Point	Machi	O	47
	Chirundu Point	Machr	O	46
	Mara Rocks	Mamr	B	58
	Nkukuti Point	Mank	O	119
	Ruarwe	Marw	B	43
Tropheops species 'olive'	Chindazowa Point	Olchi	O	61
	Chirundu Point	Olchr	O	50
	Mara Rocks	Olmr	B	53
	Mwafufu	Olmwa	O	30
	Nkukuti Point	Olnk	B	110
	Ruarwe	Olrw	B	39
	Cape Manulo	Olcm	B	29

5.2.2 *Genetic Distance Measures*

Allele frequencies at the six loci were used to calculate a statistic of genetic distance between pairs of populations. A variety of such genetic distance measures have been described, and they are based on a number of assumptions about the mechanism of microsatellite mutation. Four genetic distance measures were chosen for this study; these reflect the major categories of mutation models described in the General Introduction. The Dc (chord distance, Cavalli-Sforza and Edwards, 1967) and Dps (proportion of shared alleles, Stephens *et al*, 1992) both assume no particular model of mutation. The genetic distance measure of Nei (1972) is based on the Infinite Alleles Model originally developed for allozyme data. Delta mu (Goldstein *et al*, 1995b) assumes all loci to mutate in a stepwise manner. These four distances were calculated using the software package MICROSAT (written by E. Minch, available from http://hpgl.stanford.edu/projects/microsat/). In order to assess the statistical significance of the trees generated from the matrices of distance statistics, 1000 bootstrap replicates (see Felsenstein, 1985) were carried out using MICROSAT by resampling allelic data with replacement from each locus.

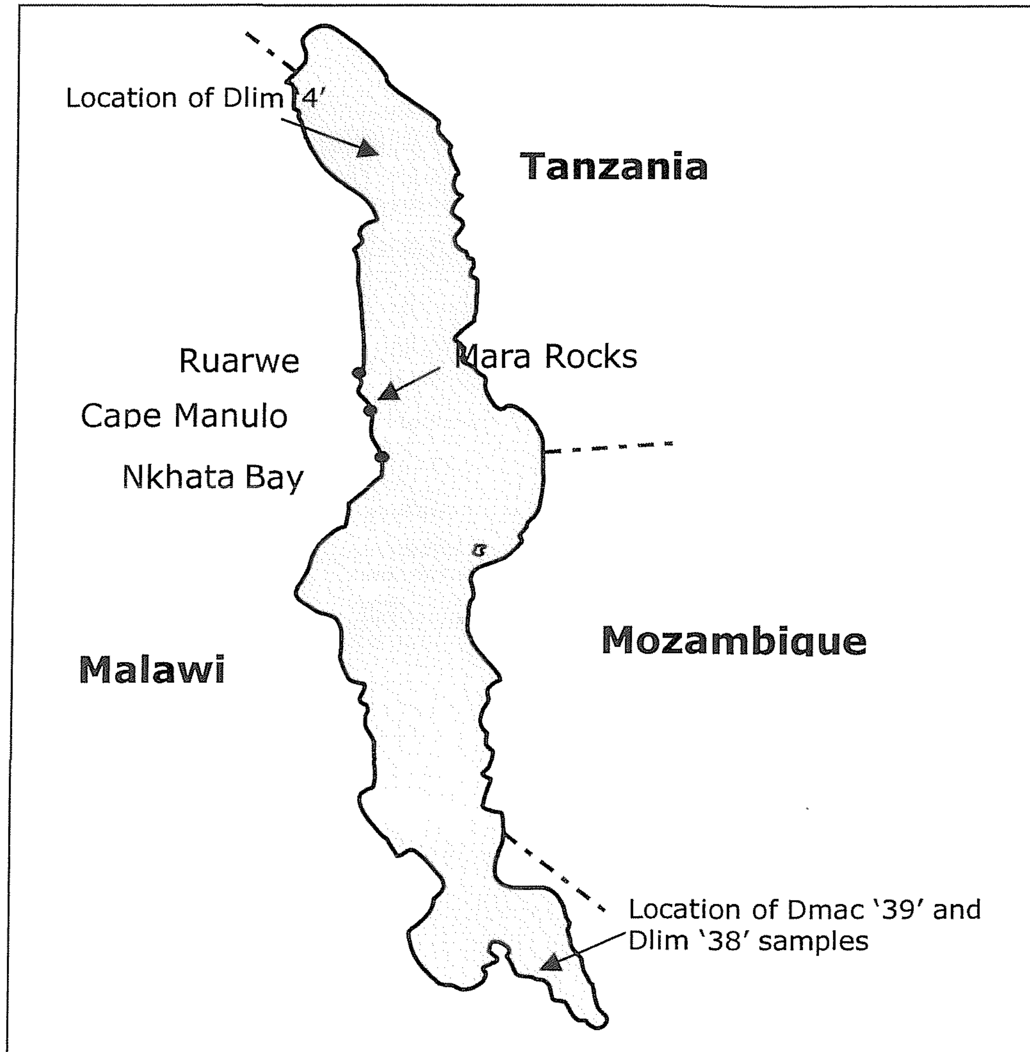
5.2.3 Tree Construction

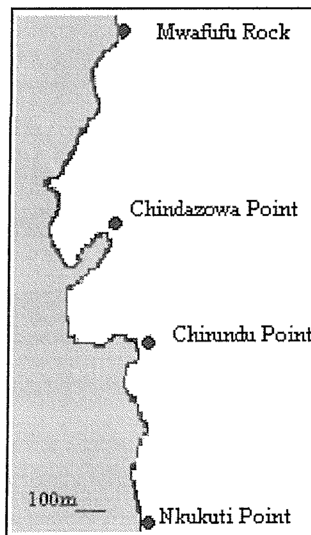
The methods available for tree construction from distance data are somewhat limited at the current time. Cluster analyses such as UPGMA can be applied to matrices of genetic distances but these produce ultrametric trees; it is assumed that all lineages have diverged equal amounts (Swofford *et al*, 1996). Neighbour-Joining (Saitou and Nei, 1987) has been shown to be applicable under a wide range of conditions (Takezaki and Nei, 1996) and does not assume equal divergence among taxa (Swofford *et al*, 1996), and was therefore chosen for this study. Trees were constructed from each of the 1000 distance matrices produced by MICROSAT for each distance statistic using the NEIGHBOR program included in the PHYLIP v3.6 software package (Felsenstein, 1993). A majority rule consensus tree was then produced using the program CONSENSE, also included in the PHYLIP software package. The consensus tree for each distance measure was drawn using TreeView v1.6.5 (written by R. Page, available from http://taxonomt.zoology.gla.ac.uk/rod/rod.html). In each tree, the *D. macrops* population was selected as an outgroup as suggested by previous molecular evidence (Figure 5.1) and differences in its ecology and morphology. Only one population was required as an outgroup, therefore the two *D. limnothrissa* populations were omitted from this part of the analysis.

5.2.4 PCA

Principal Components Analysis can be used to assess similarities between populations without attempting to infer ancestral relationships. The software package PCAGEN (written by J. Goudet, available from http://www.unil.ch/izea/softwares/pcagen.html) was used to determine the number of significant components of variation within the data for all 36 populations and to plot ordinations of the significant components.

Figure 5.2 Locations of mbuna populations sampled. Lower map shows relative locations of populations within Nkhata Bay.





5.3 Results

The loci used were highly variable both within and among populations. In the 36 populations analysed 21, 12, 48, 53, 66 and 27 different alleles were detected at the loci Pzeb4, Pzeb5, UNH002, Pzeb2, Pzeb1 and Pzeb3 respectively. Even though each locus was originally isolated from *M. zebra*, they all amplify well in different genera within the mbuna and also in both species of the more distantly related *Diplotaxodon*.

There is detectable phylogenetic structure present within the data set, although many of the nodes in the resulting trees (see Figures 5.3 to 5.6), particularly the deeper nodes have no statistical support. The tree topology appears to be very much dependent on the genetic distance statistic employed. However, some groups of conspecific populations do cluster together consistently. These groups include the *M. zebra* populations sampled from the vicinity of Nkhata Bay, The *Tropheops* 'mauve' populations from Nkhata Bay, the *Tropheops* 'olive' group as a whole, and the *M. callainos* group from Nkhata Bay.

The distance measures based on particular models of microsatellite mutation (delta mu and Nei's distance) performed relatively poorly in that they produced trees that clustered many non conspecific and non sympatric populations (Figures 5.3 and 5.4). The non mutation model dependent measures of Dps and Dc were successful in clustering conspecific populations, mostly to the exclusion of others (Figures 5.4 and 5.5). In terms of bootstrap support, the tree based on the delta mu distance had the least support with a mean support for each node of 35.9%. The trees based on Nei's distance and Dps had similar mean bootstrap values of 40.4 and 40.6% respectively. The best supported tree in terms of the mean bootstrap for each node was based on the Dc distance measure (47.6%).

It was noted that *M. zebra* species 'gold' from Nkukuti Point, *M. callainos* from Mara Rocks and *Tropheops* species 'band' from Ruarwe consistently failed to group together with conspecifics. MICROSAT was used to examine numbers of alleles and heterozygosities at each of the six loci in these populations in an attempt to explain

their unexpected positions in the phylogeny. These data are shown in Table 5.2, together with those from the nearest neighbouring population sampled. It can be seen that three loci (Pzeb1, Pzeb4 and Pzeb5) appear to be markedly less variable both in terms of numbers of alleles and heterozygosity in the Nkukuti Point population of *M. zebra* species 'gold' than in the Ruarwe population. Allele frequency distributions for both these populations can be seen in figure 5.7. The two *Tropheops* species 'band' populations also differed in heterozygosity at three loci. A comparison of the Mara Rocks *M. callainos* population with the Cape Manulo one does not reveal such immediately apparent differences as can be seen in table 5.2. There were no major visual differences in allele frequency distributions (data not shown).

The PCA revealed that the variation in the data was partitioned into three significant components that accounted for 30%, 25% and 12% of the variation respectively. A plot of the first two of these components is shown in Figure 5.9 and reveals two very distinct clusters, one of *M. callainos* and the other of *Tropheops* species 'olive'. Other clusters are formed by *M. zebra* populations and *Tropheops* species 'mauve' populations from the Nkhata Bay area. Both *M. callainos* from Mara Rocks and *M. zebra* species 'gold' do not cluster with conspecifics in this ordination. The three *Diplotaxodon* populations do not form a group distinct from the mbuna. An unrooted representation of the tree shown in Figure 5.6 reveals that while the *D. macrops* population does not cluster within any of the main mbuna clades, it has a shorter branch length than expected for such an obviously different species (Figure 5.8).

Table 5.2. Heterozygosities and number of alleles present at each of the six loci. GONK and GORW represent Nkukuti Point and Ruarwe *M. zebra* species 'gold, COBCM and COBMR represent Cape Manulo and Mara Rocks populations of *M. callainos*. The Nkukuti Point and Ruarwe populations of *Tropheops* 'band' are represented by BANDNK and BANDRW. Sample size for each population is shown in brackets.

	Population	Pzeb4	Pzeb5	UNH002	Pzeb2	Pzeb1	Pzeb3
No. alleles	GONK (50)	3	4	27	19	8	4
	GORW (30)	10	6	18	22	25	6
	COBCM (65)	5	5	22	20	39	9
	COBMR (62)	7	3	21	23	34	13
	BANDNK (53)	7	4	25	25	40	7
	BANDRW (24)	6	5	22	20	26	4
H_E	GONK	0.40	0.49	0.95	0.93	0.70	0.68
	GORW	0.80	0.68	0.91	0.92	0.93	0.68
	COBCM	0.68	0.21	0.91	0.91	0.95	0.79
	COBMR	0.47	0.20	0.90	0.90	0.96	0.81
	BANDNK	0.51	0.41	0.94	0.94	0.96	0.64
	BANDRW	0.69	0.57	0.94	0.93	0.94	0.47

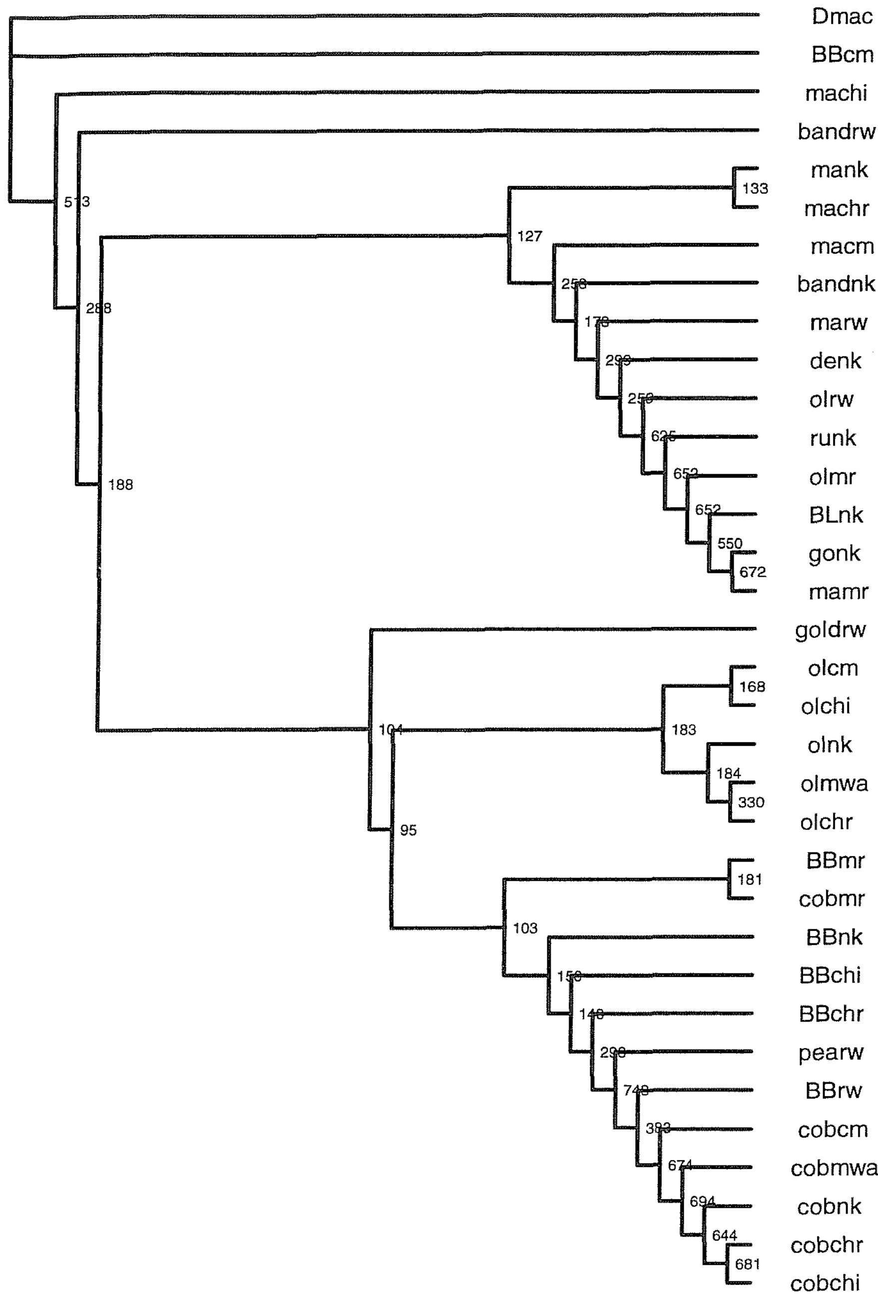


Figure 5.3. Majority rule NJ tree constructed using the delta mu distance measure of (Goldstein *et al* (1995b). Abbreviations for each population may be found in table 5.1. The figures on the nodes represent the number of times each node was recovered in 1000 bootstrap replicates.

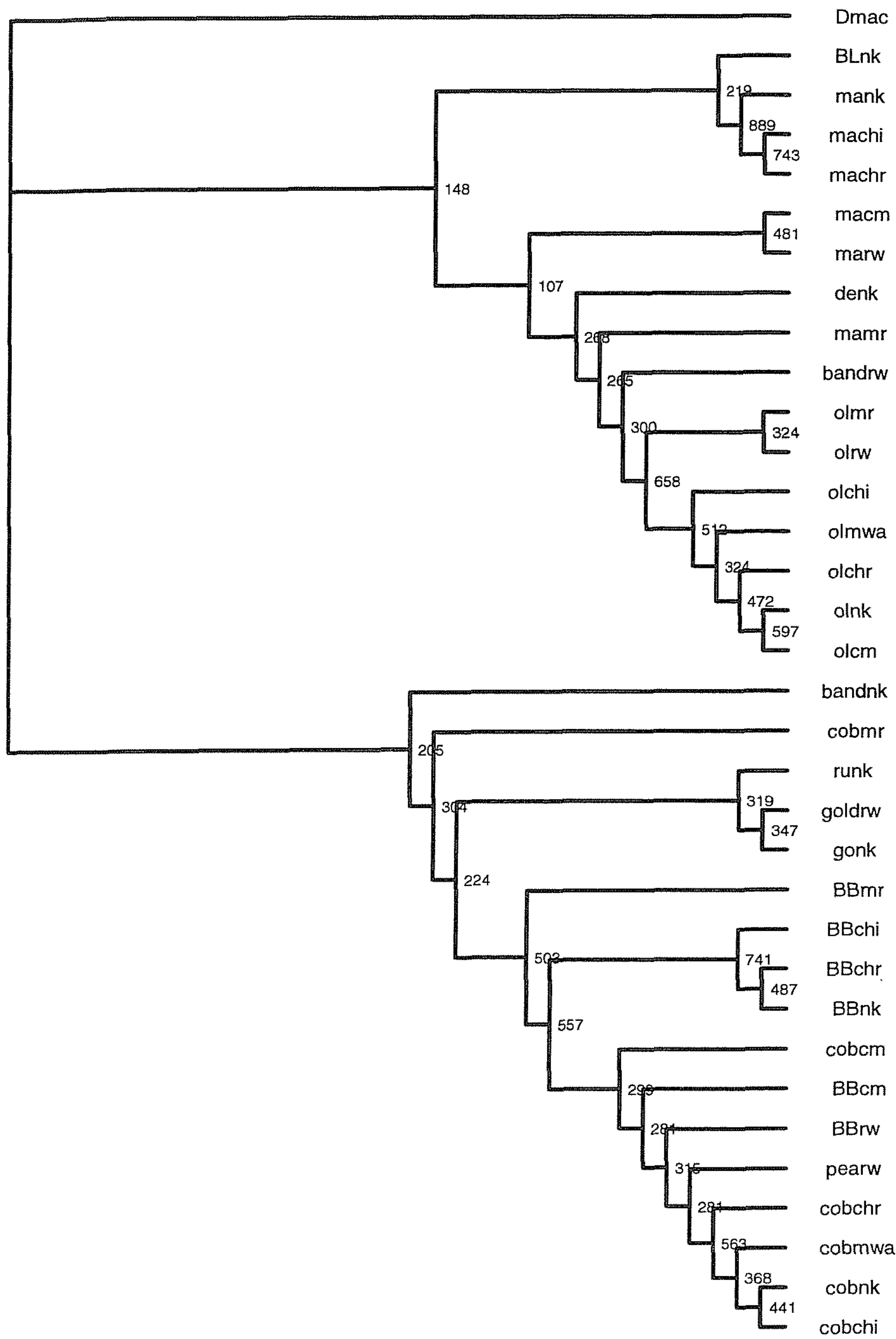


Figure 5.4. Majority rule NJ tree constructed using Nei's (1972) distance measure. Abbreviations for each population may be found in table 5.1. The figures on the nodes represent the number of times each node was recovered in 1000 bootstrap replicates.

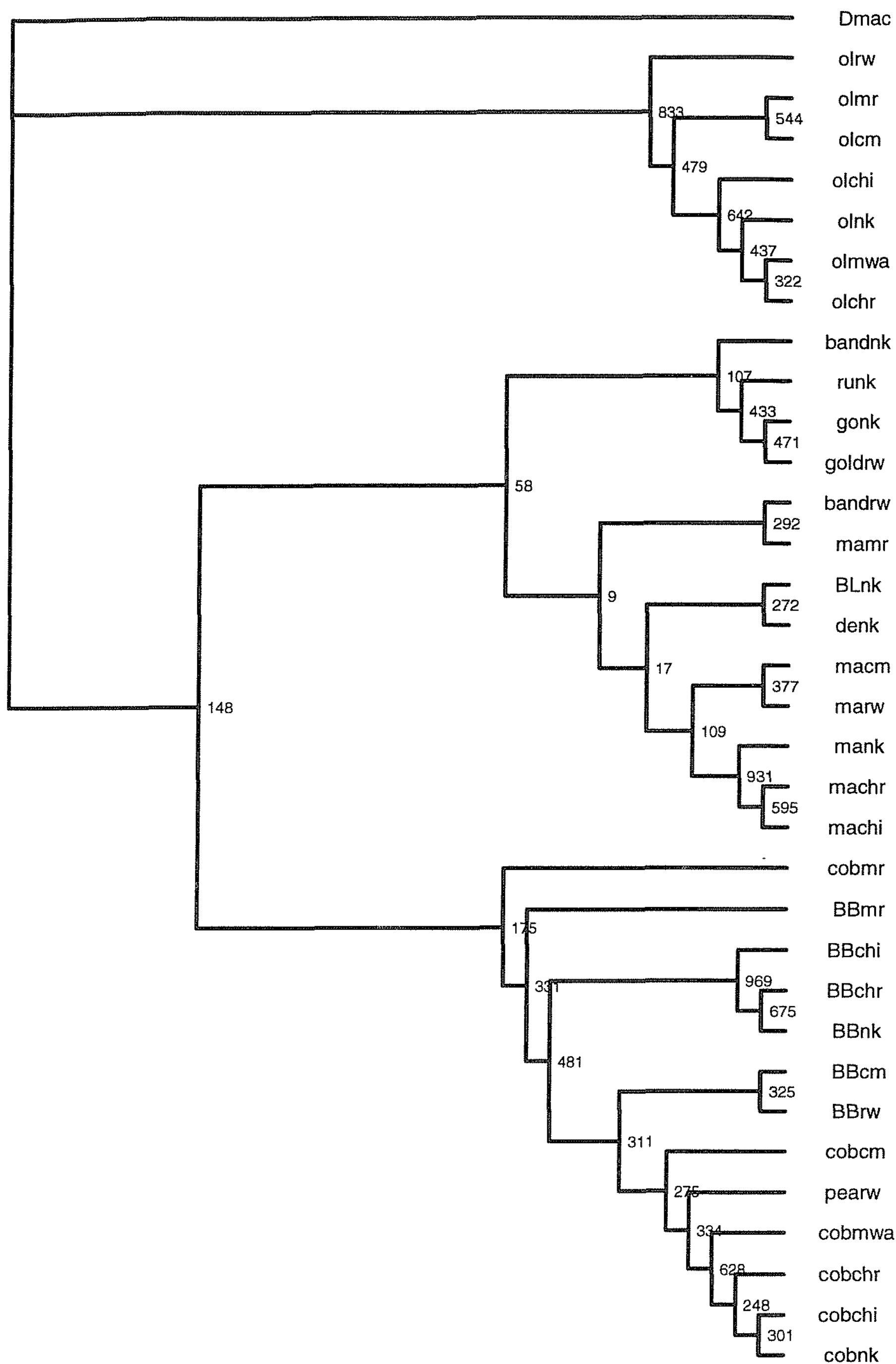


Figure 5.5. Majority rule NJ tree constructed using the Dps distance measure of Stephens *et al* (1992). Abbreviations for each population may be found in table 5.1. The figures on the nodes represent the number of times each node was recovered in 1000 bootstrap replicates.

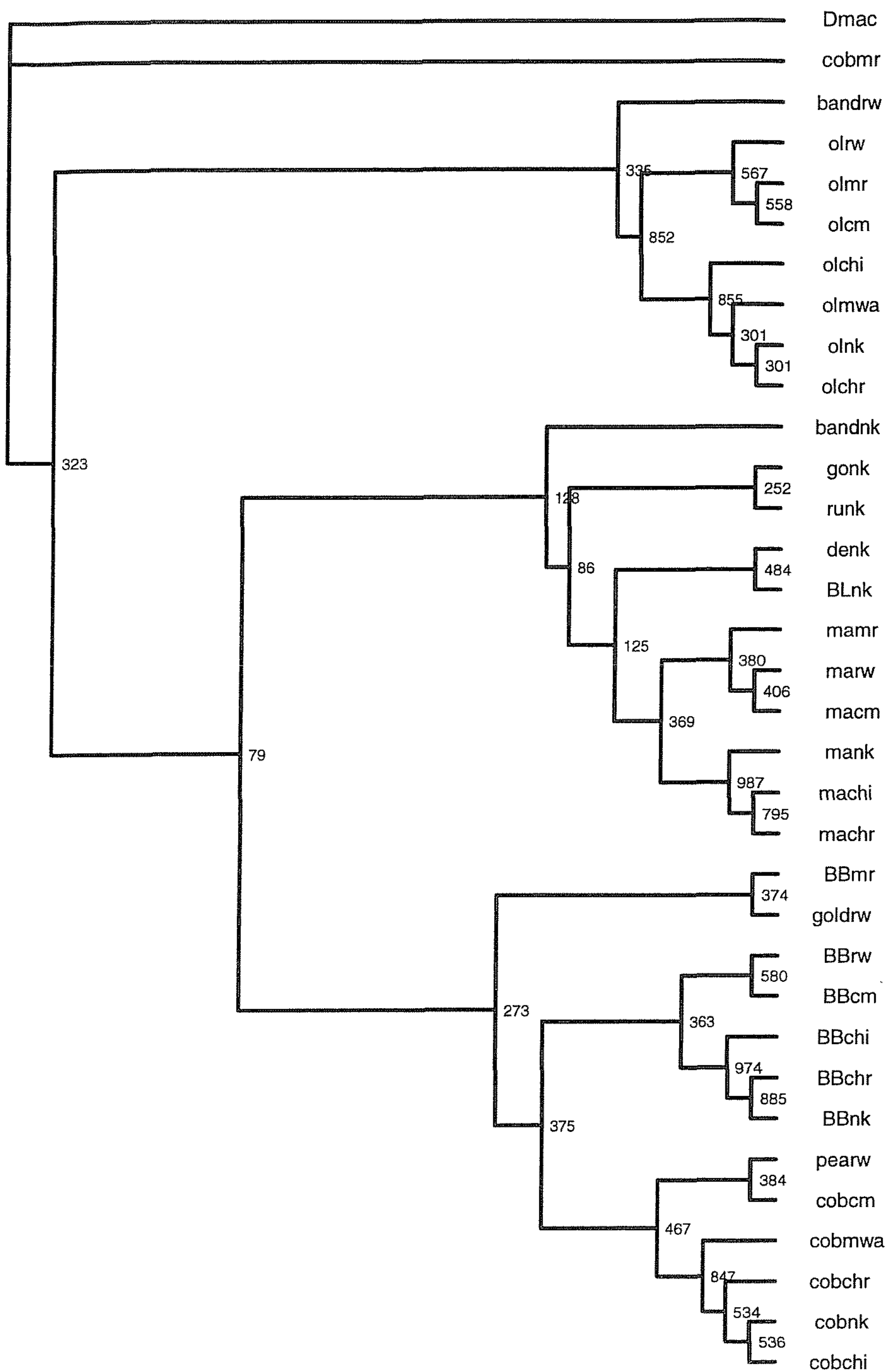


Figure 5.6. Majority rule NJ tree constructed using the Dc distance measure of Cavalli-Sforza and Edwards (1967). Abbreviations for each population may be found in table 5.1. The figures on the nodes represent the number of times each node was recovered in 1000 bootstrap replicates

Figure 5.7. Allele frequency distributions for the two populations of *M. zebra* species 'gold'. Differences in distributions between the populations can be seen at most of the loci.

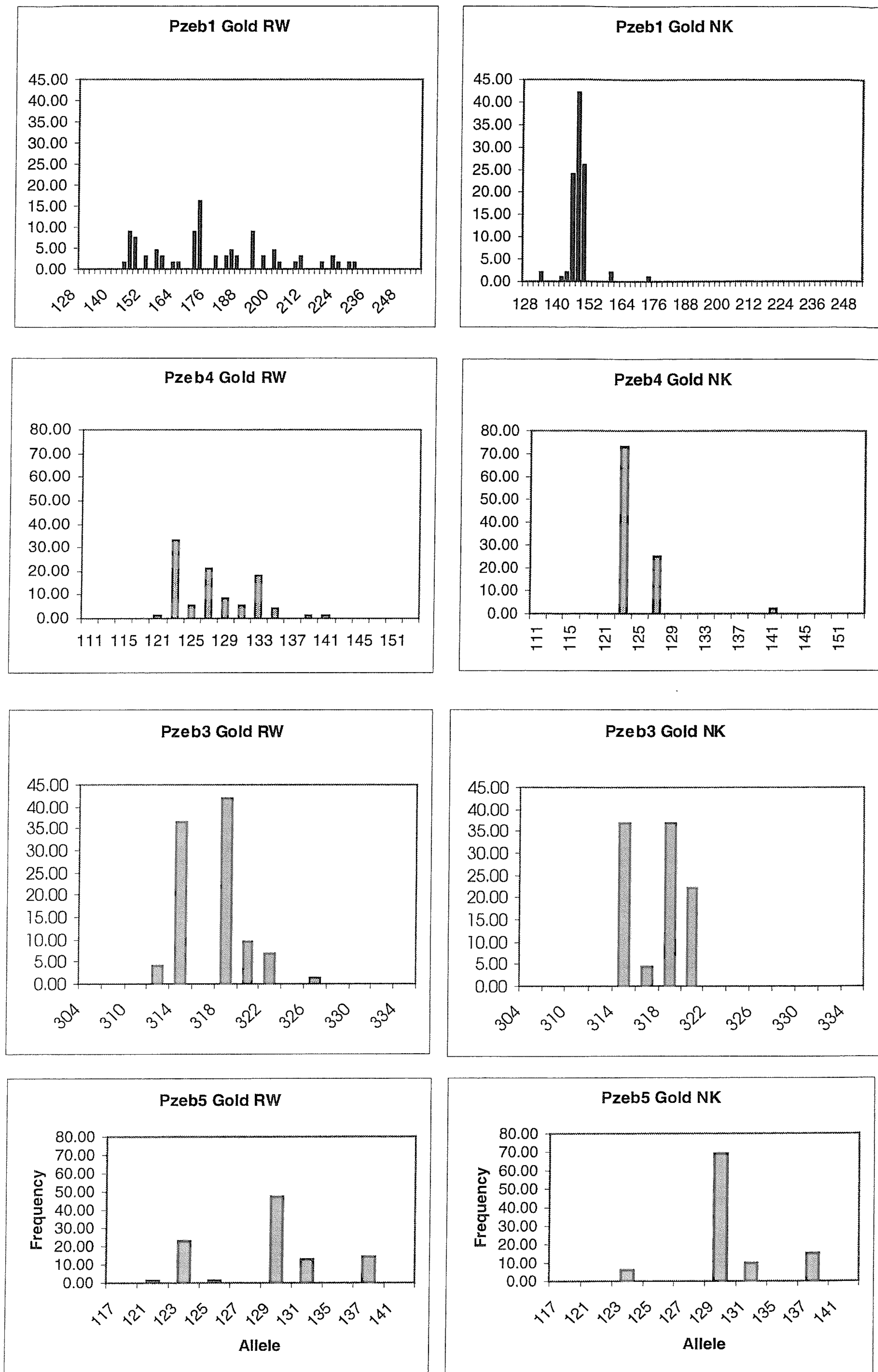


Figure 5.7 continued.

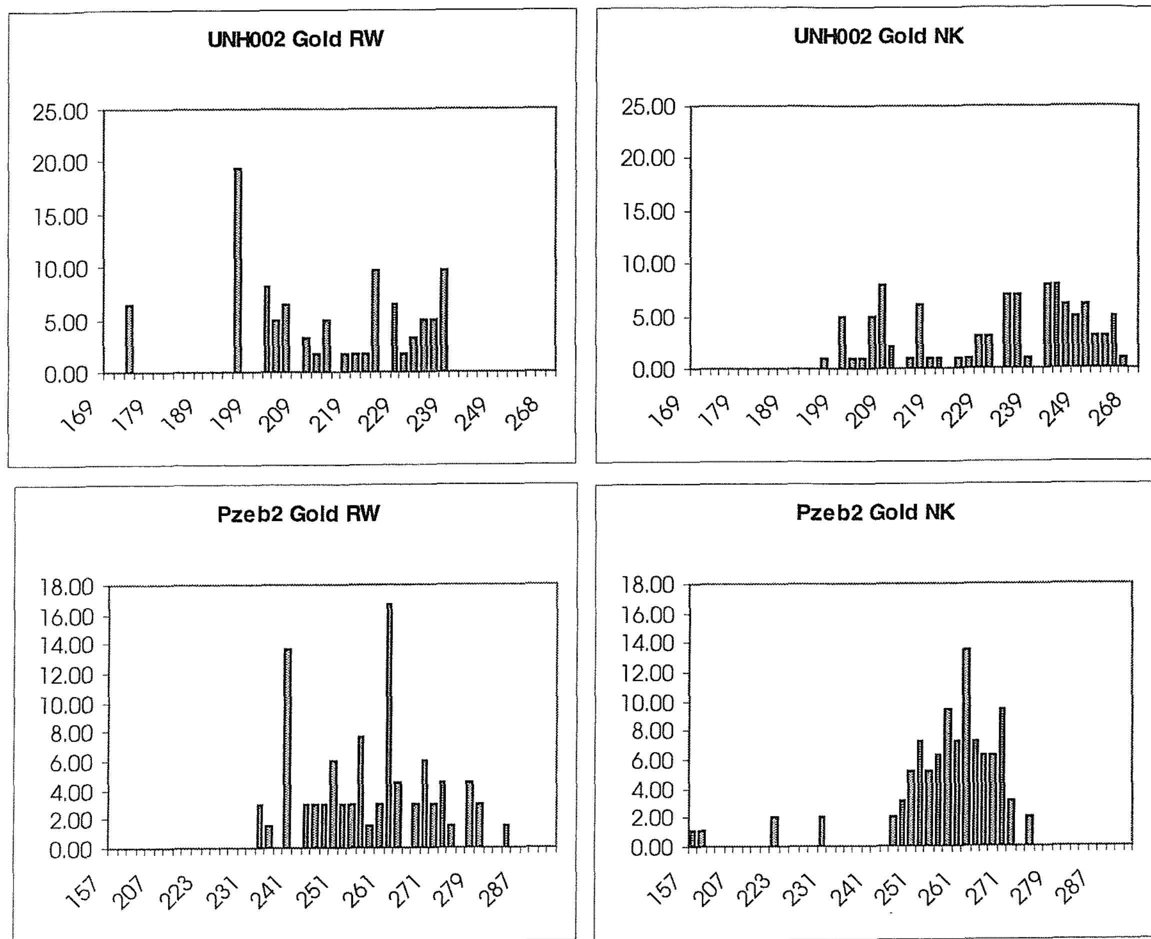
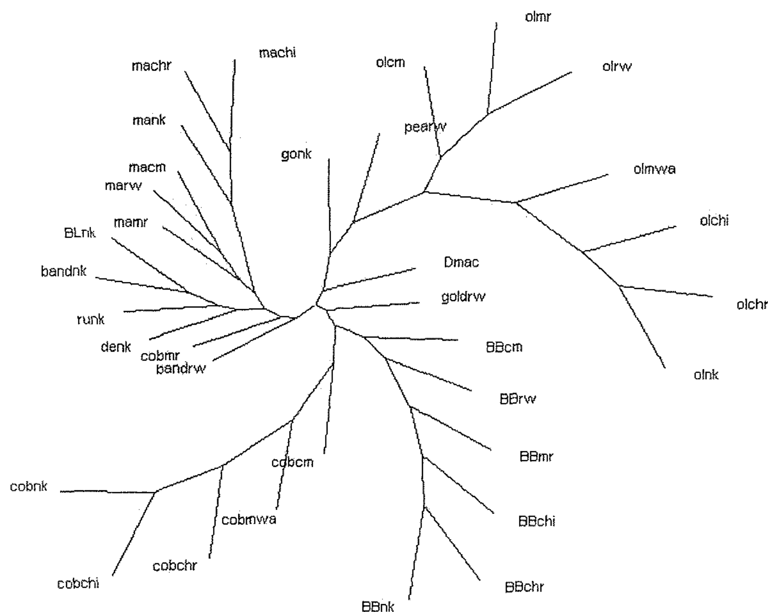


Figure 5.8. Unrooted representation of the majority rule NJ tree produced using the Dc distance measure. Population names and their abbreviation may be found in Table 5.1.



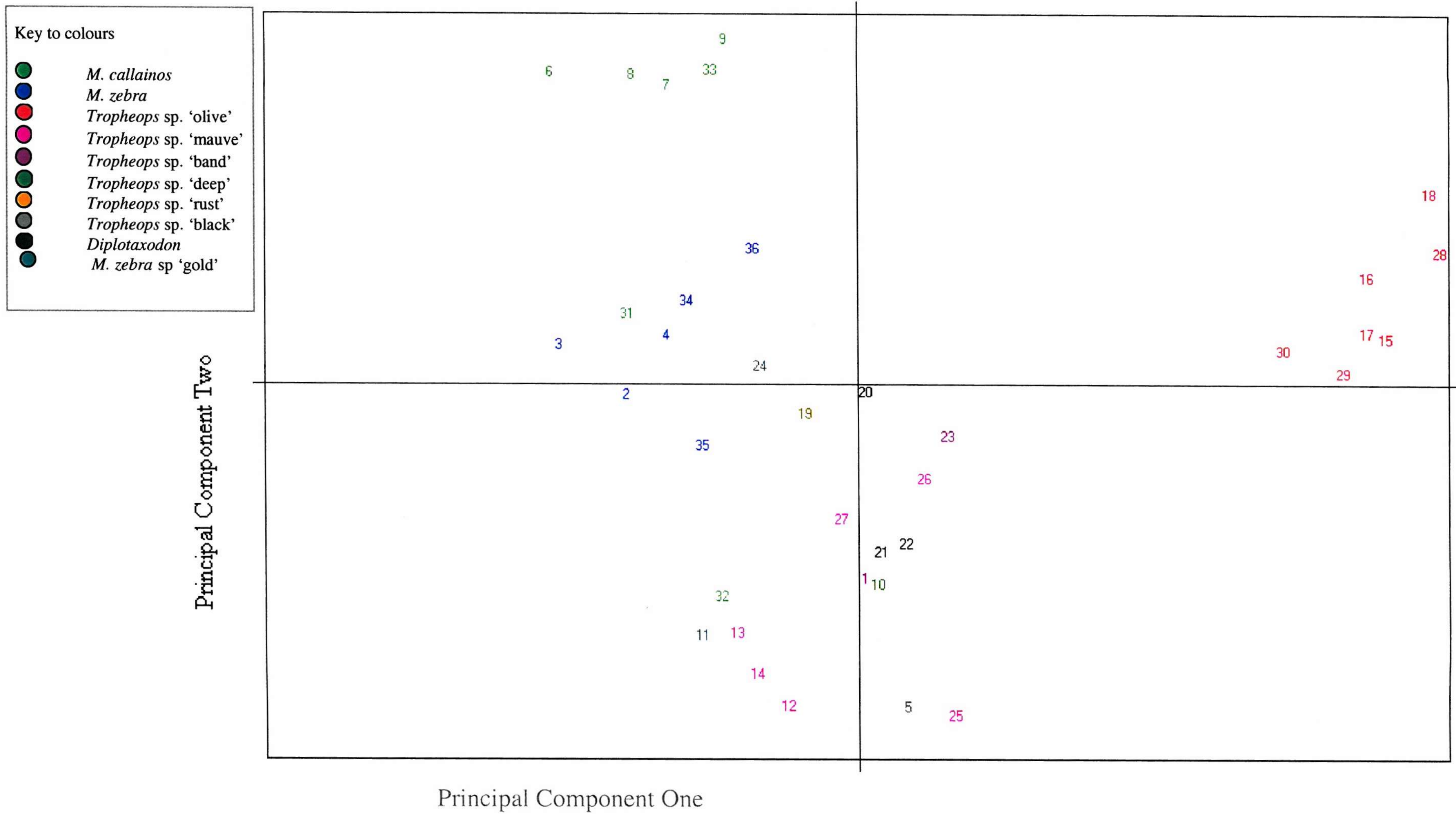


Figure 5.9. A plot of the first two principal components detected by PCAgen. These components account for 30 and 25% of the variation in the dataset respectively. Two clear clusters of populations can be easily seen; *M. callainos* (green) and *Tropheops* species 'olive' (red). The identity and location of each population may be found in Table 5.3.

Table 5.3. Species, collection location and numbers for the populations included in the ordination shown in Figure 5.8.

Species	Location	Abbreviation	Number in PCA plots	Colour in PCA plots
Tropheops sp. 'band'	NK	BANDNK	1	purple
M. zebra	CHI	BBCHI	2	blue
M. zebra	CHR	BBCHR	3	blue
M. zebra	NK	BBNK	4	blue
Tropheops sp. 'black'	NK	BLNK	5	grey
M. callainos	CHI	COBCHI	6	green
M. callainos	CHR	COBCHR	7	green
M. callainos	MWA	COBMWA	8	green
M. callainos	NK	COBNK	9	green
Tropheops sp. 'deep'	NK	DENK	10	dark green
M. zebra sp 'gold'	NK	GONK	11	aqua
Tropheops sp. 'mauve'	CHI	MACHI	12	pink
Tropheops sp. 'mauve'	CHR	MACHR	13	pink
Tropheops sp. 'mauve'	NK	MANK	14	pink
Tropheops sp. 'olive'	CHI	OLCHI	15	red
Tropheops sp. 'olive'	CHR	OLCHR	16	red
Tropheops sp. 'olive'	MWA	OLMWA	17	red
Tropheops sp. 'olive'	NK	OLNK	18	red
Tropheops sp. 'rust'	NK	RUNK	19	orange
D. limnothrissa '4'	-	DLIM4	20	black
D. limnothrissa '38'	-	DLIM38	21	black
D. macrops	-	DMAC	22	black
Tropheops sp. 'band'	RW	BANDRW	23	purple
M. zebra sp. 'gold'	RW	GOLDRW	24	aqua
Tropheops sp. 'mauve'	CM	MACM	25	pink
Tropheops sp. 'mauve'	MR	MAMR	26	pink
Tropheops sp. 'mauve'	RW	MARW	27	pink
Tropheops sp. 'olive'	CM	OLCM	28	red
Tropheops sp. 'olive'	MR	OLMR	29	red
Tropheops sp. 'olive'	RW	OLRW	30	red
M. callainos	CM	COBCM	31	green
M. callainos	MR	COBMR	32	green
M. callainos 'pearly'	RW	PEARW	33	green
M. zebra	CM	BBCM	34	blue
M. zebra	MR	BBMR	35	blue
M. zebra	RW	BBRW	36	blue

5.4 Discussion

5.4.1 Microsatellites and Phylogenetic Signal

The trees in Figures 5.3 to 5.6 show that there is some level of phylogenetic signal apparent in the data sets because there is biologically meaningful information in the tree structure. Conspecific populations tend to fall into clusters to the exclusion of others (for example, *Tropheops* species 'olive' and the *Metriaclima* species both group together). However the bootstrap values at most of the nodes are extremely low, and hence there is very little supporting evidence for any particular tree structure. Inter-specific and inter-generic relationships in particular are not well resolved. The different genetic distance statistics used produced variable results. Delta mu produced the tree with the least biological meaning, with clades consisting of populations of different species (Figure 5.3). The Dc and Dps statistics performed reasonably well, and produced broadly similar trees with three clades consisting of the *Metriaclima* populations and the 'olive' and 'mauve' *Tropheops* populations. That the tree topology is very dependent on the genetic distance statistic employed is not surprising. Each genetic distance measure has its own assumptions, not only regarding models of microsatellite mutation but also about parameters such as evolutionary rates among populations and population size. One would therefore not expect them to behave similarly for a given dataset. The mutation model dependent measures, particularly delta mu do not appear to be suitable for further use with this dataset. This may be due to the fact that two of the six loci used in this study have a complex repeat tract (Table 2.2, Chapter 2) and are thus extremely unlikely to mutate in a stepwise manner. Non-stepwise mutation has been clearly demonstrated to occur at locus Pzeb4 by van Oppen *et al* (2000). Similarly the Infinite Alleles Model assumed by the distance of Nei (1972) may be unsuitable to predict the patterns of evolution of the loci used in this study. The mixture of repeat tract types (perfect and complex) means that it is unlikely that any single mutation model will adequately describe patterns of change at each locus. This may explain why Dps and Dc, which do not make any assumptions about modes of mutation, produce trees with the highest support in terms of bootstrap values and which make the most biological sense (clustering of conspecific populations).

5.4.2 'Outlying' Populations

It is hard to talk about aberrant populations when dealing with such ill-defined trees, but three populations do not consistently cluster with conspecifics. The *M. zebra* species 'gold' population from Nkukuti Point ("gonk") almost never clustered in the *Metriaclima* clade, instead appearing in the *Tropheops* groups (Figures 5.3 and 5.6). Only using the Dps and Nei's (1972) distance statistics (Figures 5.4 and 5.5) does this population pair with its conspecific one from Ruarwe ("gorw"). In order to investigate the reason for this, heterozygosities and allele frequency distributions at all six loci were compared between the two populations (see Figure 5.7 and Table 5.2). It is immediately apparent that there are major differences in heterozygosity at loci Pzeb4, Pzeb5 and Pzeb1. In addition, there are obvious differences in allele frequency distribution, particularly at Pzeb1, Pzeb2, Pzeb4 and UNH002. The odd frequency distribution at Pzeb 1 for the 'gonk' population had been noted previously by van Oppen *et al* (1997) who hypothesised a 'selective sweep' had occurred at that locus. The authors ruled out a population bottleneck as the other five loci had allele numbers and size ranges comparable with the other mbuna populations from that location. However, when the two *M. zebra* species 'gold' populations are compared, there are major differences at more than one locus. Locus UNH002 in the 'gonk' population also appears to have had its size range severely restricted at some point in the past due to the narrow size range present. The lower diversity and heterozygosity of the 'gonk' population compared to the 'gorw' is not due to sample size differences – the 'gonk' sample size was 50 individuals, compared to the 30 individuals making up the 'gorw' sample. It is possible that demographic changes such as bottlenecks may have influenced allele sizes and frequencies in the 'gonk' population, as there are striking differences of both in more than one locus. In contrast, the *M. callainos* population from Mara Rocks also rarely clustered with conspecifics. However when allele frequency distributions (not shown) and heterozygosities (Table 5.2) were compared with the neighbouring population from Cape Manulo, striking differences as found in the *M. zebra* species 'gold' comparison were not found. In this case it is difficult to hypothesise a reason for this population's failure to cluster with conspecifics. Mara Rocks provides an extremely isolated habitat patch – migration to and from neighbouring populations is probably extremely limited. Bouteillon (1998) found smaller Fst values between the *M. callainos* populations of the Nkhata Bay area than

the northerly populations. This is not surprising considering the greater geographic distances and habitat discontinuities between Cape Manulo, Ruarwe and Mara Rocks. Therefore one might expect such populations to cluster less consistently with conspecifics. The failure of the two *Tropheops* 'band' populations to cluster in all of the trees produced was also investigated by examining the number of alleles and heterozygosities at each locus (Table 5.2). However, the sample sizes of these two populations were very different (53 from Nkukuti Point and 24 from Ruarwe). This difference arose due to the exclusion of samples from the Ruarwe population that were not typed at all six loci. The major differences in the number of alleles present at Pzeb1 and Pzeb2 could thus be at least partially explained by differences in sample size. The reason for the differences in heterozygosity is less clear. It should be noted that Ribbink *et al* (1983) classified the *Tropheops* 'band' population at Ruarwe as a different species, the results of this study indicate that this may be correct.

5.4.3 Position of Pelagic cichlids

It is notable that throughout the analyses carried out, particularly the PCA, the *Diplotaxodon* populations did not form a obvious outgroup to the mbuna species (Figure 5.9). Instead, the three *Diplotaxodon* populations clustered within the mbuna. A second PCA omitting the *M. callainos* and *Tropheops* 'olive' clusters also failed to cluster the *Diplotaxodon* populations distinctly from the remaining mbuna populations data not shown. The four trees shown in Figures 5.3 to 5.6 are drawn having manually identified *D. macrops* as an outgroup to the mbuna populations. This is justified based on other evidence such as the mtDNA tree of Idid *et al* (unpublished, Figure 5.1) which shows *Diplotaxodon* to form a separate, distinct clade from the mbuna. The data provided by allele frequencies at six microsatellite loci do not result in the same structure. It is unlikely that *Diplotaxodon* is recently diverged from an mbuna ancestor as the two groups have a vastly different morphology and ecology. It is possible that the six locus data set is unable to provide the resolution to examine relationships between the mbuna species and more distant groups such as the pelagic cichlids. However this result may indicate the extreme rapidity of lineage divergence that occurred in Lake Malawi, and thus the relative genetic similarity of the major cichlid clades within the lake.

5.4.4 Problems with Microsatellite Phylogenies

Although microsatellites have been successfully used to estimate the phylogeny of several groups of closely related species, their use in the reconstruction of evolutionary histories is not straightforward. Microsatellites have not been used to study the phylogenetics of more distantly related organisms. One possible reason for this is that phylogenetic signal is lost over longer periods of time due to the high mutation rate of such stretches of repetitive DNA. There also appears to be an upper and lower boundary to the number of repeats per allele. Obviously, the number of repeats cannot drop below zero without the microsatellite locus disappearing altogether. Alleles with small numbers of repeats also seem to be less susceptible to mutation in terms of loss or addition of repeats (Taylor *et al*, 1999). An upper boundary is likely to be present simply because alleles with extremely large numbers of repeats are never observed (Goldstein and Pollock, 1997), and because at certain loci, abnormally large expanses of simple tandem repeats are deleterious (see review by Kashi and Soller, 1999). The existence of an undefined upper threshold to the number of repeats possible at a particular locus is of even greater importance in the light of some published studies which report that microsatellite mutation is biased towards increasing rather than decreasing repeat number (Goldstein and Pollock, 1997). It has also been noted that large alleles are more mutable than smaller ones. (Jin *et al*, 1996)

The nature of the repeat region plays an important factor in determining mutation rates. Interruptions to the repeat region appear to stabilise it, preventing further mutational strand slippage (Taylor *et al*, 1999). The base composition of the repeat motif also influences mutation rate. Bactrog *et al* (2000) found that GT/CA repeats have the highest mutation rates, and AT/TA the lowest. There is also some suggestion that tetranucleotides have a higher mutation rate than dinucleotides (Weber and Wong, 1993). The loci used in the present study are of a variety of types, from perfect uninterrupted repeats to interrupted and extremely complex repeat tracts. It is unlikely that a single genetic distance statistic could reflect the mutation mechanisms and rates occurring in all six loci. This is perhaps why the statistics which make no assumption about mutation models (Dc and Dps) have performed better than delta mu, which

assumes the SMM. Paektau *et al* (1997) found similar results when testing various distance measures. The authors attributed this to the higher variance of stepwise mutation based measures which makes them less useful to study short periods of evolutionary time. Interestingly they also found no statistic could distinguish between close sister species and a species which diverged from them a long period of time before they split. Other studies have suggested that simply measuring allele sizes within a population overlooks a great deal of other sequence changes, both within the repeat and the flanking regions (Orti, Pearse and Avise, 1997; Angers and Bernatchez, 1997).

For any of these reasons, it is possible that the phylogenetic signal regarding the relationship between *Diplotaxodon* and the mbuna has been diminished over evolutionary time. The high mutation rate of microsatellite loci combined with restrictions on allelic divergence (upper and lower boundaries to the number of repeat units) means that only a limited amount of phylogenetic distance can be represented by allele frequencies. To test this, it would be useful to include another outgroup in the analysis external to both the mbuna and *Diplotaxodon*. Such an outgroup would ideally be a species not present in the Lake Malawi flock. An alternative explanation for the relative similarity of the mbuna and the *Diplotaxodon* species used is that the manual correction of data collected from different sequencers could have been faulty, and real differences between the allele frequencies of the mbuna and pelagic cichlids erased.

Estoup and Angers (1998) have suggested that robust phylogenies may be estimated from microsatellite data provided that all loci used match certain criteria. Loci with high numbers of repeats should not be included as these tend to have mutation rates that are too high. Loci with interrupted and complex repeat tracts are also deemed unsuitable as it is unlikely that mutation patterns will follow the stepwise mutation model. The authors suggest that at least 25 loci with less than 20 repeat units that show a uniform evolutionary rate across the relevant taxa are required for phylogenetic analysis of distantly related populations. For the analysis of closely related populations and species the authors suggested that locus type was not as important as locus number. However, it is obvious that isolating and testing the required loci would be an extremely lengthy procedure.

5.4.5 Conclusions

Despite the potential problems described above, the present study has indicated the phylogenetic potential of microsatellites, particularly when combined with a non-mutation model dependent distance statistic such as D_c. Six loci is recognised as an extremely small number with which to attempt phylogenetic reconstruction. A tree with higher levels of bootstrap support might be found if the number of loci used was doubled, or even tripled. However, with around fifty to one hundred individuals per population to type, this would represent a highly labour and cost intensive exercise. Microsatellites might therefore be more useful when analysing small groups of closely related populations or species in order to prevent costs becoming prohibitive. In this context, microsatellites could be used to study relationships within a particular species group such as *Tropheops* 'mauve' or 'olive' or the *Metriaclima* species within the mbuna.

5.5 REFERENCES

Angers B and Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution* 14 : 230-238

Bachtrog D, Agis M, Imhof M and Schlotterer C (2000) Microsatellite variability differs between dinucleotide repeat motifs – evidence from *Drosophila melanogaster*. *Mol. Biol. Evol.* 17 : 1277-1285

Bouteillon P (1998) Evolution and Speciation of Cichlid Fishes in Lake Malawi. MPhil thesis, University of East Anglia.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR and Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368 : 455-457

Cavalli-Sforza LL and Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics* 19 : 233-257

Estoup A and Angers B (1998) Microsatellites and minisatellites for molecular ecology : theoretical and empirical considerations. In *Advances in Molecular Ecology* ed GR Carvalho. IOS Press.

Feldman MW, Bergman A, Pollock DD, and Golstein DB (1997) Microsatellite genetic distances with range constraints : analytic description and problems of estimation. *Genetics* 145 : 207-216

Felsenstein J (1985) Confidence limits on phylogenies, an approach using the bootstrap. *Evolution* 783-791

Felsenstein J (1993) PHYLIP (Phylogeney Inference Package) v3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Goldstein DB and Pollock DD (1997) Launching microsatellites : a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 88 : 335-342

Goldstein DB, Linares AR, Cavalli-Sforza LL and Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139 : 463-471

Goldstein DB, Linares AR, Cavalli-Sforza LL and Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the USA* 92: 6723-6727

Harr B, Weiss S, David JR, Brem G and Schlotterer C (1998) A microsatellite based multilocus phylogeny of the Drosophila melanogaster species complex. *Current Biology* 8 : 1183-1186

Jin L, Macaubas C, Hallmayer J, Kimura A and Mignot E (1996) Mutation rate varies amongst alleles at a microsatellite locus: phylogenetic evidence. *Proc. Natl. Acad. Sci. USA* 93 : 15285-15288

Kashi Y and Soller M (1999) Functional roles of microsatellites and mini satellites. In *Microsatellites – Evolution and Applications*. eds DB Goldstein and C Schlotterer. Oxford University Press

Kornfield I and Parker A (1997) Molecular systematics of a rapidly evolving species flock : the mbuna of Lake Malawi and the search for phylogenetic signal. In *Molecular Systematics of Fishes* ed T Kocher. Academic Press.

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution* 8 : 279-284

Meyer A, Kocher TD, Basasibwaki P and Wilson AC (1990) Monophyletic Origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347 : 550-553

Moran P and Kornfield I (1993) Retention of an ancestral polymorphism in the mbuna species flock of Lake Malawi. *Molecular Biology and Evolution* 10: 1015-1029

Nagel S, Tichy H, Mayer W, Takeszaki N, Takhhata N and Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proceedings of the Royal Society of London B* 267:1049-1061

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

Nishida N (1991) Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes ; inferences from allozyme data. *Experientia* 47 ; 974-979

Orti G, Pearse DE and Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences USA* 94 : 10745-10749

Paektau D, Waits LP, Clarkson PL, Craighead L and Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147 : 1943-1957

Petren K, Grant BR and Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proc. R. Soc. Lond. B* 266 : 321-329

Sage RD, Loiselle PV, Basasibwaki and Wilson AC (1984) Molecular versus morphological change among cichlid fishes of Lake Victoria. In *Evolution of Fish Species Flocks* eds AA Eschelle and I Kornfield. University of Maine Press.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425

Shaw PW, Turner GF, Idid MR, Robinson RL and Carvalho GR (2000) Genetic populations structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proc. R. Soc. Lond. B* 267 : 2273-2280

Streelman JT, Zardoya R, Meyer A and Karl SA (1998) Multilocus phylogeny of cichlid fishes (Pisces : Perciformes) : evolutionary comparison of microsatellite and single copy nuclear loci. *Molecular Biology and Evolution* 15 : 798-808

Stephens JC, Gilbert DA, Yuhki N and O'Brian SJ (1992) Estimation of heterozygosity for single probe multilocus DNA fingerprints. *Molecular Biology and Evolution* 9: 729-743

Takezaki N and Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144 389-399

Taylor JS, Durkin, JMH and Breden F (1999) The death of a microsatellite: a perspective on microsatellite interruptions. *Mol. Biol. Evol.* 16 : 567-572

van Oppen MJH, Turner GF, Rico C, Deutsch JC, Ibrahim KM, Robinson RL and Hewitt GM (1997) Unusually fine scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London Series B* 264 : 1803-1812.

Weber JL and Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2 : 1123-1128

5.6 Chapter 5 Appendix

Details of genetic distance statistics:

Nei (1972)

$$D = -\ln \frac{\sum_m \sum_i P_{1mi} P_{2mi}}{(\sum \sum P_{1mi}^2)^{1/2} (\sum \sum P_{2mi}^2)^{1/2}}$$

Where m is summed over loci, i is summed over alleles and P_{1mi} is the frequency of the ith allele at the mth locus in population 1.

Cavialli-Sfoza and Edwards (1967) Chord Distance

$$D^2 = \frac{4 \sum_m (1 - \sum P_{1mi}^{1/2} P_{2mi}^{1/2})}{\sum_m (a_m - 1)}$$

Where m indexes loci, I is summed over alleles at the mth locus and a is the number of alleles at the mth locus.

Delta mu squared (Goldstein *et al*, 1995)

$$D_{dmu} = (mu(A) - mu(B))^2$$

Where mu (A) is the mean allele size for population A

Proportion of shared alleles

$$D_{ps} = \frac{\sum_i MIN(P(A(i)), P(B(i)))}{n}$$

D_{ps} is defined as the mean of the minima of the relative frequencies of alleles in the taxa being compared. N represents the total number of alleles for all loci.

6 Testing parallel evolution of colour patterns in the mbuna using AFLPs

6.1 Introduction

One outstanding attribute of the mbuna cichlids of Lake Malawi is the enormous variety of colour patterns present within this group. Much of this variation occurs in male coloration, which is thought to form the basis for female mate choice in non pair forming East African cichlids (e.g. Knight *et al*, 1998; Seehausen *et al*, 1997). These colour patterns evolve rapidly; Owen *et al* (1990) suggest that the southern end of the lake was dry as little as 250 years ago, which could mean that the endemic colour forms present in this area have evolved since the subsequent lake level rise. Many morphologically and ecologically similar species pairs differ only in male coloration (Ribbink *et al*, 1983), suggesting that divergence of male colour frequently occurs with speciation events (Seehausen *et al*, 1999a). Differences in male colour patterns form the basis of the current intra-generic taxonomy in the mbuna (Ribbink *et al*, 1983). However, there is a degree of variation in colour pattern between allopatric populations of the same putative species, and the current lack of potential for geneflow between these populations makes assigning specific status difficult. A few examples of clines in colour pattern have also been recorded (Arnegard *et al*, 1999; Ribbink *et al*, 1983). Deutsch (1997) found there was more variability in colour hue between members of a species complex (a group of similar species within a genus) than between species complexes, and also noted that similar colours may be found within different complexes and even genera. Seehausen *et al* (1999b) hypothesised that this repeated occurrence of similar coloration across genera was the result of disruptive sexual selection. The fact that colours appear to evolve rapidly coupled with the very real ecological and morphological differences that exist between genera suggest that colour patterns found in more than one genus/species complex are likely to have evolved more than once in parallel. Parallel evolution has been defined as 'the evolution of similar or identical features independently in related lineages' (Futuyama, 1986 p554). However, in order to test hypotheses of colour pattern evolution in Lake Malawi a robust phylogeny is required. This has proved difficult to obtain even after the advent of DNA sequencing and molecular markers due to a lack of variation an incomplete lineage

sorting in the genes examined (Meyer, 1993). The extremely short period of evolutionary time over which mbuna genera and species complexes have diverged means that many molecular markers do not mutate at a high enough rate to reflect divergence events. Rapidly mutating markers such as microsatellites may prove applicable to this problem as they appear to be ideal for reconstructing the evolutionary history of recently diverged groups (see previous chapter and references therein). However, determining evolutionary relationships amongst large numbers of populations/species represents a time consuming and costly exercise as large numbers of individuals need to be typed at many loci (Estoup and Angers, 1998). This certainly seems to be the case for the mbuna, as discussed in the previous chapter. AFLPs (the name is derived from the PCR amplification of restriction fragment length polymorphisms) are a relatively new class of molecular marker originally developed by Vos *et al* (1995). A brief summary of the methods behind AFLPs is given in the General Introduction to this thesis. AFLPs have already been used in a number of phylogenetic studies of closely related organisms (Sharbel *et al*, 2000; Wilding *et al*; 2001; Giannasi *et al*, 2001; Parsons and Shaw, 2001). In the first investigation to utilise AFLPs on African cichlids Albertson *et al* (1999) examined speciation and trophic divergence within mbuna genera collected from sites at the southern end of Lake Malawi. They were able to determine relationships within and between the four genera studied and although intrageneric branch lengths were short, the resulting tree was statistically well supported. AFLPs are therefore a class of marker that appears to be suitable for phylogenetic reconstruction and hypothesis testing. This study aims to use AFLPs to provide a robust phylogeny of mbuna species from a variety of geographic locations within the lake. This phylogeny will be used to test the hypotheses of parallel evolution of four colour traits. By testing species from a wide geographic range it will also provide indications of the phylogenetic levels (intra/inter specific and generic) that are best resolved by AFLP data. This will provide valuable information for the planning and execution of other similar studies on the mbuna.

6.2 Methods

6.2.1 Traits and species used

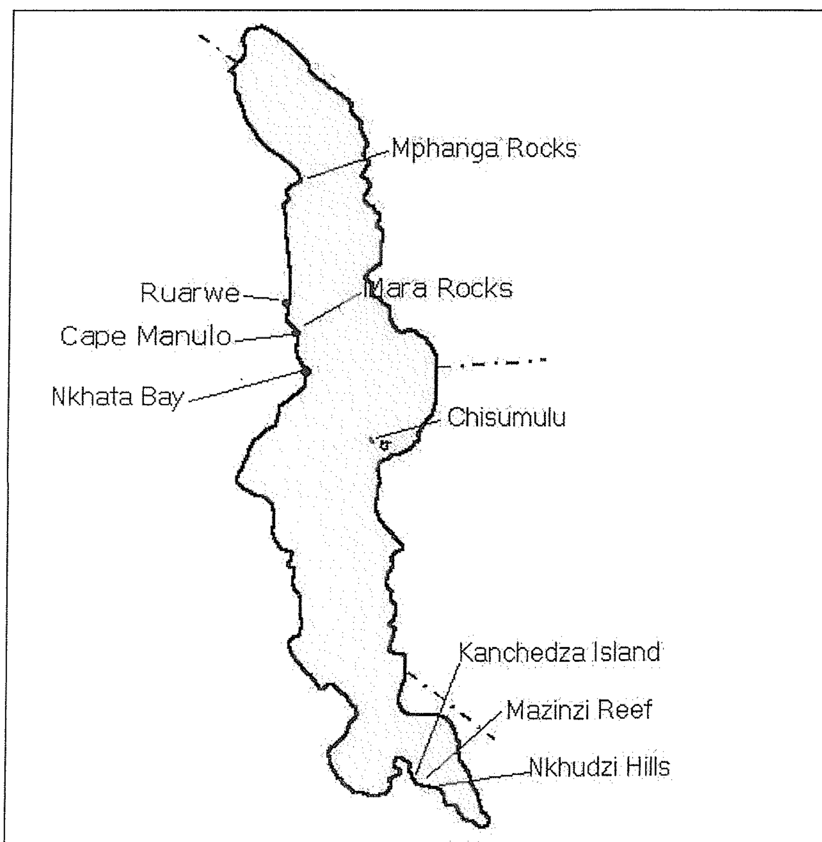
Five colour traits were selected for investigation. Of these traits, four were male based (yellow chest, red dorsal fin, yellow ground colouration and blue ground coloration) and one was female based (bright white or orange coloration). The female trait is more complex as many mbuna species or populations are polymorphic for female colour; whilst most individuals have a cryptic coloration with dark vertical bars, a minority exhibit a bright orange or a blotched coloration. However, in some species, the cryptic form with vertical bars is absent, and it is this lack of a barred form which is being tested. Table 6.1 shows the original collection locale and presence or absence of the trait for each species used in the study. The locations of collection sites are also indicated in Figure 6.2. Three representatives of each species/population were used. Several outgroups were included to test the recovery of deeper phylogenetic relationships. These were *Oreochromis niloticus*, four species of cichlids from Lake Victoria (see Table 6.1) and *Rhamphochromis* sp (a lake Malawi non-mbuna). The total number of samples in the study was limited to 64 as this was the maximum number that could be run on a single gel on the ABI 377 sequencer (this avoids problems associated with scoring bands across different gels).

Table 6.1 Presence and absence of the traits tested and geographic locations for the populations used in this study. Photographs or representatives from each population may be seen in Figure 6.2. With regard to white/orange females, + indicates that the population is fixed for the trait, -/+ indicates the trait is polymorphic in the population. ? indicates unknown whether white/orange trait occurs in this population.* *M. zebra* (Mara Rocks and Ruarwe) is also known as *M. fainzillberi*

Species Used	Collection locale	Yellow Chest	Yellow ground coloration	Blue ground coloration	Red Dorsal fin	White /orange females
Metriaclima callainos	Nkhata Bay	-	-	+	-	+/-
M callainos	Mara Rocks	-	-	+	-	+/-
M callainos 'pearly'	Ruarwe	-	-	-	-	+
M.estherae	Mozambique	-	-	-	-	+
M. zebra 'gold'	Nkhata Bay	+	+	-	-	-
Tropheops species'olive'	Nkhata Bay	+	+	-	-	-
M. benetos	Mazinzi Reef	-	-	+	-	-
M. zebra	Chisumulu	-	-	+	-	?
M. zebra	Cape Manulo	-	-	+	-	?
M. zebra	Mazinzi Reef	-	-	+	-	?
Tropheops 'mauve blue'	Ruarwe	-	-	+	-	-
Labeotropheus fuelliborni	Lab stock	-	-	+	-	-/+

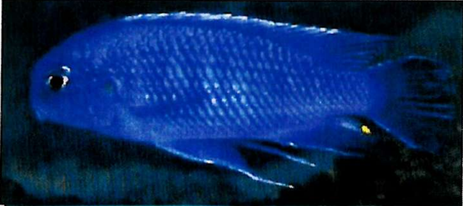
Species Used	Collection locale	Yellow Chest	Yellow ground coloration	Blue ground coloration	Red Dorsal fin	White /orange females
M. emmiltos	Mphanga Rocks	-	-	+	+	-
M. sandaracinos	Nkhudzi hills	-	-	+	+	-
M. aurora	?	+	-	+	-	-
*M. zebra**	Mara Rocks	+	-	+	-	?
*M. zebra**	Ruarwe	+	-	+	-	?
M. xanstomachus	Kanchedza Island	+	-	+	-	-
Pseudotropheus elongatus	Mara Rocks	+	+	-	-	-
Tropheops 'mauve aurora'	Ruarwe	+	-	+	-	-
Oreochromis niloticus	Lab stock					
Rhamphochromis sp	Lake Malawi					
Neochromis omnicaeruleus	Lake Victoria					
Neochromis species unicuspid scraper	Lake Victoria	Outgroups				
Pundamilia pundamilia	Lake Victoria					
P. nyererei	Lake Victoria					

Figure 6.1. Sampling locations for populations used in this study.



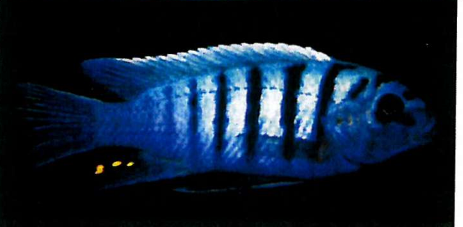
b) Blue ground coloration (male) – other than seen in a), c) and d)

M. zebra – Cape Manulo[1]

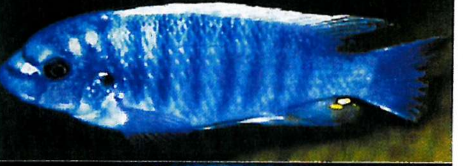


Tropheops 'mauve blue'[1]

M. zebra Chisumulu[1]



M. zebra Mazinzi Reef[1]



Labeotropheus fuelleborni[1]

M. benetos[1]

b) Red dorsal fin (male)

M. emmiltos[1]

M. sandaracinos[1]

c) yellow ground coloration (male)

M. zebra 'gold'[1]

Tropheops 'olive'[1]

d) Females lacking vertical bars. This trait also occurs in females of *Labeotropheus fuelleborni*, for which no photo was available.

M.estherae[2]
Left – female
Right male

M. callainos 'pearly'[1]
Left – female
Right - male

M. callainos[2]
Nkhata Bay**
Right - male
Left – female (blue morph)

d) cont. ** The Nkhata Bay population of *M. callainos* is polymorphic in terms of female coloration, with individuals exhibiting blue, white and blotched coloration. A similar situation exists for *M. callainos* at Mara Rocks.

6.2.2 Sample collection and extraction.

The samples were collected in the form of clips from the soft-rayed posterior portion of the dorsal fin and were preserved in absolute ethanol until required for extraction. A few samples (*M. zebra* from Cape Manulo, *M. callainos* from Mara rocks and *M. xanstomachus*) were obtained as DNA extracts. DNA was extracted from fin clips using the proteinase K/chloroform protocol of Rico *et al* (1992, see chapter 2 for details) and resuspended in 200 μl TE buffer (100mM Tris, 10mM EDTA). The concentration of the DNA extracts was assessed with a TKO 100 DNA Mini-Fluorometer (Hoefer Scientific Instruments) using calf thymus DNA to obtain a standard curve. DNA extracts were diluted to obtain a 10ng/μl solution of each sample.

6.2.3 AFLP Procedure

The original protocol of Vos *et al* (1995) was followed, with some modifications suggested by RC Albertson (pers. com.). All oligonucleotides were manufactured by Sigma-Genosys Ltd, and all PCR programs were carried out using a Hybaid Touchdown thermocycler. The AFLP protocol consists of four stages; restriction/ligation, pre-selective amplification, selective amplification and visualisation.

6.2.3.1 Restriction/ligation

The restriction enzymes selected were *EcoR*I and *Mse*I, as used in Albertson *et al* (1999). The restriction and ligation of samples was carried out at the same time – a departure from the protocol of Vos *et al* (1995) who separated these two stages. An enzyme master mix was prepared; this comprised 0.1µl T4 DNA ligase buffer (300mM Tris-HCl, 100mM $MgCl_2$, 100mM DTT, 10mM ATP), 0.1µl 0.5M NaCl, 50 ng BSA, 1 unit *Mse*I (New England Biolabs), 1 unit *EcoR*I (Promega) and 1 unit T4 DNA ligase (Promega) per sample. The restriction ligation mix comprised 1 µl T4 DNA ligase buffer with ATP, 1µl 0.5M NaCl, 0.5 µg BSA, 1µl enzyme master mix as described above, 5 pmol of *EcoR*I adapters, 50 pmol *Mse*I adapters and 55ng DNA. The resiction/ligation mixture was then incubated in a water bath for 2 hours at 37°C. After incubation, the restriction/ligation mixture was diluted with 190µl 0.1x TE buffer.

6.2.3.2 Pre-selective amplification

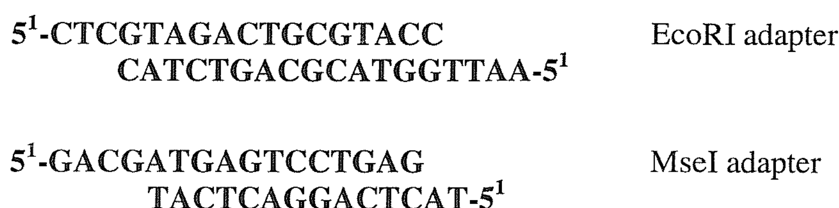
This stage is required to reduce the number of fragments amplified from complex genomes and to reduce primer binding mismatches during selective PCR (Vos *et al*, 1995). It acts to improve repeatability and ease of interpretation of AFLP profiles on gels. Pre-selective PCR was carried out in 20 µl volumes containing Thermoprime Plus PCR buffer (ABgene; 75mM Tris-HCl, 20mM $(NH_4)SO_4$, 0.1% Tween 20), 0.2mM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of each of the pre-selective *Mse*I and *EcoR*I primers (see Table 6.2) and 0.4 units of Thermoprime Plus polymerase (ABgene). The pre-selective PCR program consisted of 20 cycles of 30s at 94°C, 1 minute at 56°C and 1 minute at 72 °C. The PCR products were then diluted with 180µl 0.1x TE buffer for use as a template in the selective PCR.

6.2.3.3 Selective Amplification

In total, 18 different pairs of selective primers were used (see Table 6.3). Each *EcoR*I primer was labelled at 5^1 with FAM or HEX to allow detection by the ABI 377 sequencer. The PCR parameters were identical to the pre-selective parameters except that only 3 µl of the diluted pre-selective products were included, and 1 and 5 pmol of the *Eco*RI primer and the *Mse*I primer respectively were added instead of the pre-selective primers. The selective PCR required a relatively complex thermal profile. The

first 13 cycles consisted of 30s at 94°C, 30s at the annealing temperature and 1 minute at 72°c. The annealing temperature started at 65°C and was reduced by 0.7°C per cycle. The subsequent 23 cycles were of 30s at 94°C, 30s at 56°C and 1 min at 72 °C. 7 µl of the reaction products were run on a 1% agarose gel containing ethidium bromide to ensure the PCR had worked.

Figure 6.3 Sequence of the adapters ligated to the cut DNA fragments. Single stranded oligonucleotides were PAGE purified after manufacture. Adapter pairs were diluted to the required concentration (5 µM for EcoRI adapters and 50 µm for MseI adapters) and heated to 70°C. After 10 minutes the adapter solutions were allowed to cool to room temperature in order to anneal.



5^1-CTCGTAGACTGCGTACC
CATCTGACGCATGGTTAA-5^1 EcoRI adapter

5^1-GACGATGAGTCCTGAG
TACTCAGGACTCAT-5^1 MseI adapter

Table 6.2 Sequence structure of the PCR primers used in this study. The core sequence is complementary to that of the relevant adapter molecule.

	Core Sequence	Enzyme Specific	Preselective	Selective
EcoRI	5^1-GACTGCGTACC	AATTC	A	XX-3^1
MseI	5^1-GATGAGTCCTGAG	TAA	C	XX-3^1

Table 6.3 The combinations of selective primers used in this study. The names of the primers reflect the final three bases at the 3^1 end of the primer and the name of the enzyme. Primer pairs 2 and 5 were discarded from the study as they were found to be identical to primer pairs 12 and 16. The dye labels attached to the EcoRI primers are indicated and the primers were HPLC purified after manufacture.

	EACA (FAM)	EAGG (FAM)	EACT (FAM)	EAGC (FAM)	EACC (FAM)	EAAC (HEX)
MCAC	1				19	
MCAT		8		6		
MCAA	16		3	7	18	
MCTA	20		14		9	
MCAG			15	17		11
MCTT	4	12				
MCTG		13				10

6.2.4 Visualisation

Band visualisation was achieved using an ABI 377 sequencer upgraded to recognise 64 lanes and the GeneScan data collection software package (both by Applied Biosystems). 1 μl of the selective amplification products was combined with 1.1μl loading buffer (0.16 μl size standard, 0.26 μl loading dye [50mg/ml blue dextran; 25mM EDTA], 0.68μl deionised formamide). The combined mixture was denatured at 90°C for 2 minutes and then loaded onto the sequencer via 64 lane membrane combs using the protocol of Toonen (submitted). The internal size standard used was GeneScan-500 ROX (Applied Biosystems). All samples amplified using a particular primer pair were run on the same gel to avoid problems of scoring bands across different gels. Gel conditions were similar to those described in Chapter 2. After electrophoresis, fragments were sized using the GeneScan software package (Applied Biosystems).

6.2.5 Band Scoring

Three different methods of band scoring were tested. The first used the Genotyper software package (Applied Biosystems) to output a list of peaks for each sample (lane of the gel). This list was transferred into a spreadsheet package to allow manual alignment of peaks between lanes. However, the peak recognition algorithms in the Genescan and Genotyper software packages is somewhat basic (Garnhardt and Kocher, submitted) and may result in extra peaks being detected due to fluctuations in the fluorescence of a fragment. Only peaks between 75 and 400 bases in size were scored as the intensity of peaks decreased markedly with fragment size.

The second method of peak scoring used the automatic binning algorithm available in Genotyper to identify peaks and assign them to categories (loci). Categories were created by first labelling all peaks on the gel between 75 and 400 bases with heights of at least 50, 150, 250, 400 or 600 relative fluorescence units (rfu). All peak heights were scaled relative to the intensity of the total signal from the first lane on each gel. This has the effect of normalising the signal across the gel to account for small differences in the amount of PCR products loaded into each lane. Peaks were then assigned to categories with a size tolerance of +/- 0.4 bases. All peaks on the gel were then

relabelled using the categories created from the orginal labels except that this time all peaks higher than 50 rfu were labelled. Thus the creation of a category depended on the presence of a peak higher than a certain threshold, but once created, all peaks within the category size range higher than 50 rfu were labelled and included. Categories that overlapped in size were excluded. Five data files were created using this method (one for all gels using each peak height threshold).

The third method involved the use of the freeware software package Genographer (© Montana State University, 1998, http://hordeum.oscs.montana.edu/genographer/), which imports the files produced by GeneScan and produces a straightened and sized gel image. This allows semi-manual scoring of bands and enables the intensity of bands to be normalised across a gel. It also allows for more accurate scoring of less intense bands and permits ambiguous areas of the gel (such as smears) to be left unscored. Using this software, peaks between 75 and 500 bases could be scored unambiguously.

In all three methods each gel was scored in a binary fashion for band presence/absence. Analysis of the data was carried out both including and excluding monomorphic bands as these may introduce bias into calculations of genetic distance (Link *et al*, 1995).

6.2.6 Testing Reproducibility

In order to test whether AFLP profiles were consistently reproducible, twelve samples were subjected to three different restriction-ligation reactions. The restriction-ligation products were then amplified using primer pair 1. PCR products were then run on the same gel under the conditions described above except that a GeneScan350 size standard was used. Peaks were scored using Genographer, and band presence/absence was compared for each sample across the three restriction-ligation reactions. The number of bands scored differently within samples was averaged to obtain the mean difference between the three restriction-ligations.

6.2.7 Phylogenetic Analysis

The data sets were tested for non-random structure using the g1 statistic (Hillis and Huelsenbeck, 1992) because even random data may produce an apparently well-supported tree. This test was carried out by using the software package PAUP* vb4.01 to generate 100000 random trees from the AFLP data and to calculate the g1 statistic based on the distribution of their lengths. The g1 values were compared to empirically determined critical values (Hillis and Huelsenbeck, 1992) to check for statistically significant non-random structure within the data set. The data were tested for non-random structure at different levels - with and without outgroups and well supported clades. Tests were also carried out with only one out of the three representatives of each mbuna taxon present. A second series of tests involved the five data files generated by scoring peaks at different thresholds using Genotyper. In this set of tests, only the mbuna populations were included in the analysis.

Phylogenetic reconstruction from AFLP data requires the calculation of genetic distance between samples and the subsequent construction of a tree from the distance matrix. Two genetic distance statistics that can be applied to AFLP data are those of Link *et al* (1995) and Nei and Li (1979) which are implemented in the software package Treecon (van de Peer and de Wachter, 1994). Equations for these distance statistics are shown below. Tree building was carried out using the neighbour-joining algorithm (Saitou and Nei, 1987) as in previously published AFLP based phylogenetic studies (Giannassi *et al*, 2001; Albertson *et al*, 1999; Parsons and Shaw, 2001). Bootstrapping (Felsenstein, 1985) was used as an indication of the statistical support for a particular tree topology. Trees were constructed using all populations for which AFLP data was available and also for just the mbuna populations using *Rhamphochromis* as an outgroup. The *Oreochromis* and Victorian samples were omitted in an attempt to better resolve relationships between the *Metriaclima* populations.

Distance of Link *et al* (1995)

$$D_{Link} = \frac{N_x + N_y}{N_x + N_y + N_{xy}}$$

Where N_x is the number of bands present in sample x but not in sample y, N_y is the number of bands present in sample y but not in x, and N_{xy} is the number of bands shared by x and y

Distance of Nei and Li (1979)

$$D_{NeiLi} = \frac{2N_{xy}}{N_x + N_y}$$

Where N_{xy} is the number of bands present in sample x and y, N_x is the number of bands present in x, and N_y is the number of bands present in y.

6.3 Results

6.3.1 Reproducibility

The AFLP profiles generated from three different restriction-ligations of twelve samples were extremely similar when run on the same gel. A total of 87 different bands were scored and these differed by only an average of 7.6% between restriction-ligations for the same sample. Only two bands were detected consistently in the second and third restriction-ligations but not in the first. If these two bands were excluded, overall reproducibility became 94%.

A summary of each data set can be found in Table 6.4. Each primer pair produced a large number of polymorphic bands, although the number of bands detected depended on the peak scoring method (Figure 6.4). Of the data files constructed using Genotyper with various peak thresholds, the file resulting from a threshold of 250 rfu returned the most negative g1 statistic based on the lengths of 100000 random tree lengths (Table 6.5). This data set was therefore chosen for comparison with the other two methods of peak scoring. However, it was noted that the data sets obtained using Genotyper with a

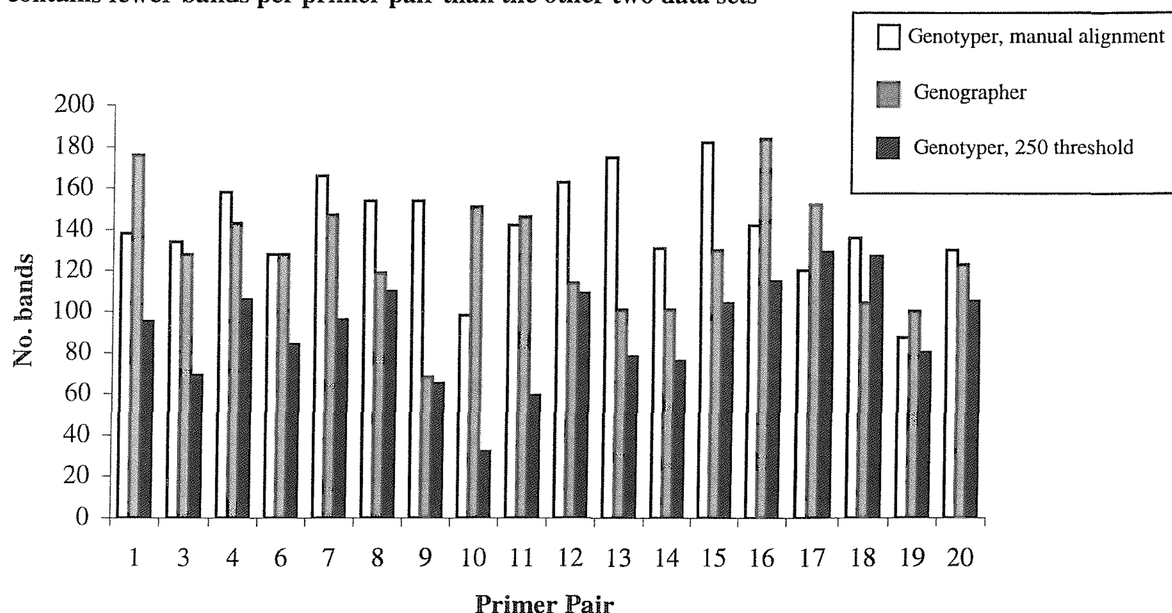
peak height threshold of 50, 150, 250 and 400 rfu produced very similar trees (see Appendix at the end of this chapter).

The percentage of invariable bands in each data set varied only slightly (4.6% using Genotyper with manual alignment, 3.4% using Genotyper [250 peak threshold] and 5.6% using Genographer). The data set produced by Genotyper with manual alignment contained the largest number of bands with a total of 2538 and had by far the highest mean number of bands per primer pair at 141. With the exception of primer pairs 10, 11, 16 and 17, the number of bands per primer pair is greater using Genotyper (manual alignment) than in either of the other two (Figure 6.4). The Genotyper (250 peak threshold) data set without exception contains the lowest number of bands, both per primer pair and in total. Trees were generated from the data sets both including and excluding monomorphic bands. No differences in overall topologies were observed. The trees shown below result from analysis of each data set including invariable bands.

Table 6.4. Comparison of the three data sets generated from 18 different AFLP primer pairs using different methods of peak scoring

	Genotyper, manual alignment	**Genotyper 250**	**Genographer**
Total number of bands	2538	1354	2315
% variable bands	96.6	95.4	94.4
Mean no. of bands per primer pair	141.0	75.2	128.61

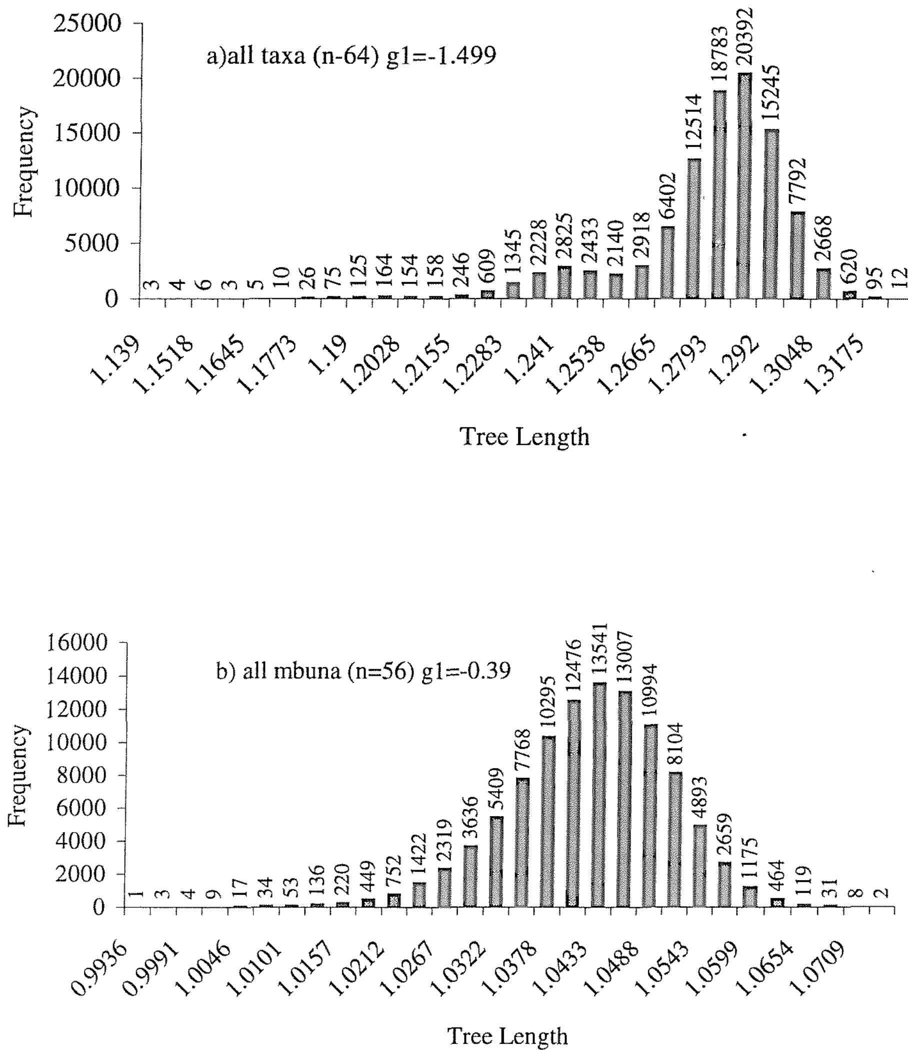
Figure 6.4. Number of bands detected per primer pair in each of the three data sets. Data set three contains fewer bands per primer pair than the other two data sets



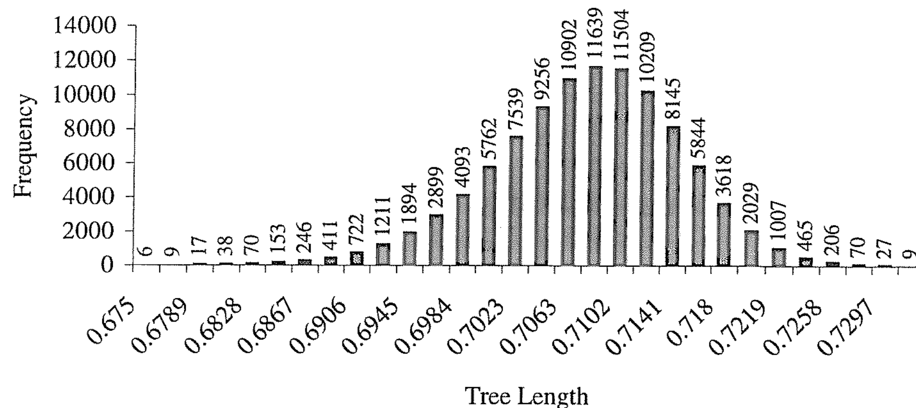
6.3.2 Non Random Structure in the data

The distributions of the lengths of 100000 random trees and the g1 statistics for each distribution can be seen in Figure 6.5 a-d. It is clear that there is significant structure to the data, as shown by the negative g1 statistics (the more negative the g1 value, the greater the degree of non-random structure present). All four sets of taxa tested contain significant non-random structure, however the degree of structure decreases as outgroups and other well-supported clades are removed. The vast majority of the structuring among the mbuna taxa appears to be due to the similarity of samples taken from the same population. When only one individual from each population is included in the analysis (Figure 6.5d) there is very little skew in the tree length distributions and the g1 value becomes very close to the critical values determined by Hillis and Huelsenbeck (1992). However, as Hillis and Huelsenbeck (1992) only determined g1 critical values for a maximum of 25 taxa and 500 characters it is difficult to accurately extrapolate critical values for an larger data set such as that used in this study.

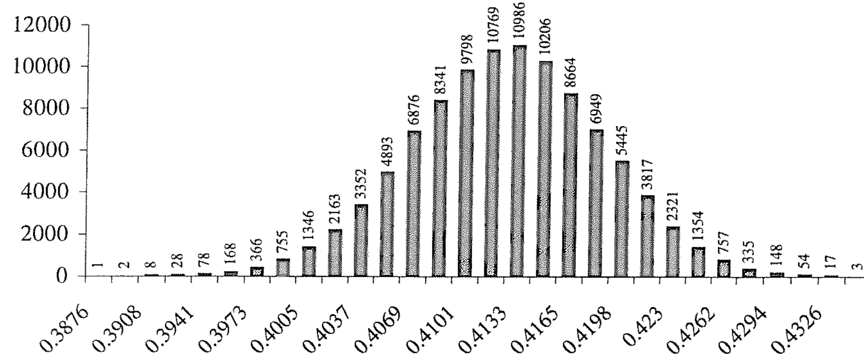
Figure 6.5. Distribution of the lengths of 100000 randomly generated trees for various numbers of populations. Peaks were scored using Genographer. a) all populations, b) all mbuna populations, c) all mbuna populations except *M. callainos/M.estherae* and *Tropheops/Labeotropheus* clades, d) as c) but with a single individual from each population. g1 statistics are indicated, g1 values for a), b) and c) are more negative than the critical values determined by Hillis and Huelsenbeck (1992) suggesting that significant non random structure is present in the data. The significance of the value for d) cannot be determined as it is difficult to extrapolate accurately from the critical values given by Hillis and Huelsenbeck (1992) for a data set of this size.



c) all mbuna except *Tropheops* and *M. callainos* clades (n=35) g1=-0.332



d) as c) but with a single individual from each taxon (n=12) g1=-0.08



A summary of the g1 statistics for the data sets scored using Genotyper with various peak height thresholds can be seen in Table 6.5. There are slight differences in g1 between the 50, 150, 250 and 400 rfu data sets but clearly at 600 rfu, the structure within the data starts to break down.

Table 6.5. Summary of the data sets resulting from scoring bands using different peak height thresholds in Genotyper. Only mbuna samples were included in the analysis.

Peak Threshold (rfu)	Total number of bands	g1
50	2012	-0.24
150	1354	-0.25
250	1639	-0.26
400	950	-0.22
600	610	-0.12

6.3.3 Changes in Peak Height Threshold

The total number of bands scored decreases dramatically as the peak height threshold in Genotyper is increased (Table 6.5). The major structure of the tree remains unchanged when the peak height threshold is changed, although bootstrap support decreases noticeably at the highest threshold (see figures 6.15 to 6.18 in the Appendix to this chapter). At a peak height threshold of 600 rfu, *Rhamphochromis* does not occupy its expected position as an outgroup to the mbuna. The trees produced by the lower peak thresholds are quite similar, despite the major differences in terms of numbers of bands scored. However it is noted that a sample of *M. zebra* from Chisumulu and *M. callainos* from Mara Rocks behave differently during analysis of these data sets. In some trees these samples cluster with other members of their population/species and in others with the *M. zebra* samples from Cape Manulo. This is also observed in the trees based on the Genotyper (250 rfu) dataset (Figures 6.9 and 6.10). Trees based on mbuna populations (excluding *Oreochromis* and the Victorian samples) are therefore not shown for this dataset, but no other major structural differences were observed compared to Figure 6.13 and 6.14.

6.3.4 'Odd' individuals

Two other individual samples were consistently noted not to cluster with conspecifics (ZebS2, a member of a southern *M. zebra* population, and Ben1, a member of the *M. benetos* population). If included in the analysis, ZebS2 became basal to the rest of the mbuna, altering the topology of the tree (see Figure 6.19 in Appendix), but only during analysis of data scored using Genographer. This sample clustered with the other two members of its population when data scored using Genotyper were analysed. Ben1 showed a tendency to cluster with the *P. elongatus* samples but only during analysis of data scored using Genotyper. Trees were generated both with and without these two samples as their aberrant position was thought to be an artefact of the way the peaks were scored; if these samples were truly as genetically divergent from other samples of the same populations then one would expect to observe this in all trees resulting from all data sets. Ideally, these 'odd' samples should be retyped to check whether they are

genuinely different from other members of their population or if an error has occurred in the AFLP process. It should also be borne in mind that the populations included in this analysis are extremely closely related (in terms of AFLP profile), which is demonstrated by the scale bars included with each of the trees. It may be that partial failure or contamination of the PCR of just one primer pair would be enough to cause these samples to cluster aberrantly.

6.3.5 Effect of the number of primer pairs used on tree resolution

Overall tree resolution was assessed using the average bootstrap support value for all nodes across a given tree. Trees were generated using successive numbers of primer pairs and both the distance measure of Link *et al* (1995) and that of Nei and Li (1979). The support for a given tree increases consistently as data from successive primer pairs are included in the analysis (Figure 6.6).

6.3.6 Tree Topology and Resolution

In all trees, all outgroup nodes were well supported with extremely high bootstrap values. The samples from Malawi and Victorian cichlids are shown to be reciprocally monophyletic. The *Rhamphochromis* sample is basal to the rest of the Malawi samples. Within the mbuna, tree topology varies slightly depending the genetic distance statistic and method of peak scoring used. In the following sections, an association of a group of populations seen in the majority of trees is regarded as a clade, especially when bootstrap support for the group is greater than 50%. It should be noted that some bootstrap levels provide at best weak statistical support for these groupings, but it is the appearance of the same groups in trees based on different methods of peak scoring or distance statistics that is of interest.

Despite topological variation due to differences in peak scoring or genetic distance statistics, the trees in Figures 6.7-6.14 show a number of major similarities. The *Tropheops* and *Labeotropheus* populations form a well supported clade in all trees with 64-100% bootstrap support. Trees generated using Genographer and Genotyper differ in the position of this clade relative to the *Metriaclima* populations. If Genographer is used, *Tropheops* forms a sister group to the rest of the mbuna populations (Figures 6.11

and 6.12). However, if Genotyper is used, then the *Tropheops* clade becomes part of a group consisting of the northern *Metriaclima* and *Pseudotropheus* populations. This topology has a bootstrap support value of 55% in one case (Figure 6.8).

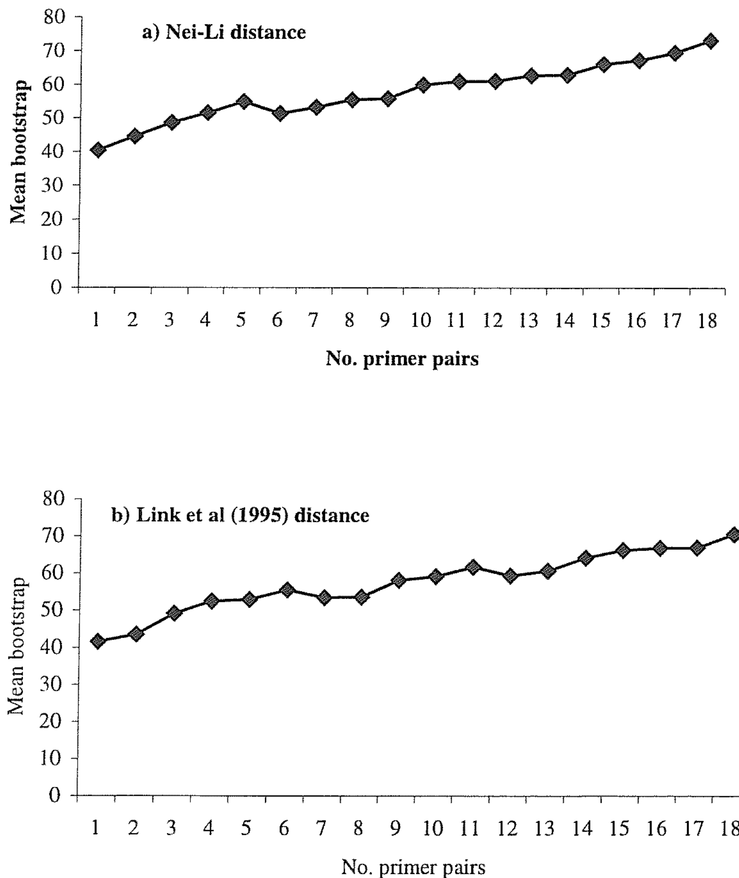


Figure 6.6. Effect of addition of successive loci on mean bootstrap value. Trees were constructed data from Genographer and Nei and Li (1979) and Link *et al* (1995) distances.

Another probable clade is formed by the *M. callainos* and *M.estherae* populations. In all trees these populations cluster together, and in three out of the eight trees the clade has >50% support. A third clade consists of the southern *Metriaclima* populations. This cluster is found in three out of the six trees constructed using all samples (Figures 6.8, 6.11 and 6.12), and is present in all of the smaller trees based on mbuna samples only, with bootstrap support of 59-75%. Only in one of the trees does any of the southern *Metriaclima* populations (*M. benetos* and *M. xanstomachus*) cluster with the northern

Metriaclima populations with >50% bootstrap support (Figure 6.9). Interestingly, this topology is not exhibited in the tree based on the mbuna subset of the same data (not shown). Figures 6.13 and 6.14 indicate a monophyletic origin of the northern *Metriaclima* and *Pseudotropheus* populations. This topology is supported by bootstrap values of 51-63%. These trees vary in the position of the *Tropheops/Labeotropheus* clade relative to *Metriaclima* populations; figure 6.13 shows it to be part of the group formed by *Metriaclima* and *Pseudotropheus* populations from the northern part of the lake whilst in figure 6.14 the *Tropheops* clade is a sister group to the group comprised of all *Metriaclima* and *Pseudotropheus* populations.

6.3.7 Population Relationships and Evolution of Colour Traits

When only mbuna populations, together with *Rhamphochromis*, were analysed, the major clades were equally or better resolved than when the *Oreochromis* and Victorian samples were included. The two trees based on data from mbuna populations (Figures 6.13 and 6.14) show a number of major similarities with regard to relationships between populations. The main difference between them is the position of the *Tropheops/Labeotropheus* clade as described above, and the fact that *M. zebra* (Cape Manulo) occupies a basal position in Figure 6.13. Trees based on the same populations and constructed using the distance measure of Link *et al* (1995) showed extremely similar structure (see figure 6.20 in Appendix).

6.3.7.1 Yellow chest

In all trees *Tropheops* species 'mauve aurora' is shown to be more closely related to the sympatric non yellow chested *Tropheops* 'mauve blue type', and also to the differently coloured *Tropheops* 'olive'. These three populations all fall into a well supported clade. The yellow chested *M. aurora* and *M. zebra* (sampled from Ruarwe and Mara Rocks) are absent from this clade, as is the yellow chested species *M. xanstomachus* from the southern part of the lake and *P. elongatus*. None of these four latter species form a clade together, although Figure 6.13 exhibits a mixed cluster of *M. zebra* (Ruarwe and Mara Rocks samples). Both trees indicate that *M. xanstomachus* consistently clusters with the other *Metriaclima* populations from the south of the lake rather than the extremely similar *M. zebra* populations from Ruarwe and Mara Rocks.

6.3.7.2 Yellow ground colouration

As mentioned above, *Tropheops* 'olive' clusters with other *Tropheops* populations and with *Labeotropheus*, despite the fact that males from these populations are mainly blue in colour. The species similar to *Tropheops* 'olive' in male colour (*M. zebra* species 'gold' does not form part of this clade. The two species with yellow ground coloration have been scored as also having the yellow chest trait, as it is impossible to determine the presence/absence of the latter. Apart from in the *Tropheops* clade, there appears to be no phylogenetic association between the yellow ground coloration and yellow chest traits.

6.3.7.3 Blue male ground coloration

This is a very widespread trait, and is seen in fifteen of the twenty mbuna populations sampled. The trait occurs in all of the clades detected (southern *Metriaclima*, *Tropheops/Labeotropheus* and *M. callainos/estherae*) and thus does not appear to be phylogenetically constrained.

6.3.7.4 Red Dorsal Fin

The two species with red dorsal fins are not shown to be sister taxa. In fact it appears that they are relatively distantly related as *M. sandaracinos* clusters consistently with the *Metriaclima* species from the south of the lake, to the exclusion of *M. emmiltos*. This indicates that the red dorsal fin trait has evolved twice in the populations sampled.

6.3.7.5 White/orange females

A distinct clade is formed by the *Metriaclima* species that exhibit bright white or orange coloration in females (the three *M. callainos* populations, and *M. estherae*). This cluster is observed in every tree generated. However, orange/white females are also found in the *Labeotropheus* population, which is more closely related to the *Tropheops* species that do not exhibit this trait. Orange/white females may also be found in some of the *Metriaclima zebra* populations sampled (detailed data currently lacking). However, in the case of *Labeotropheus*, *M. zebra* and *M. callainos* (Nkhata Bay and Mara Rocks), the trait is polymorphic.

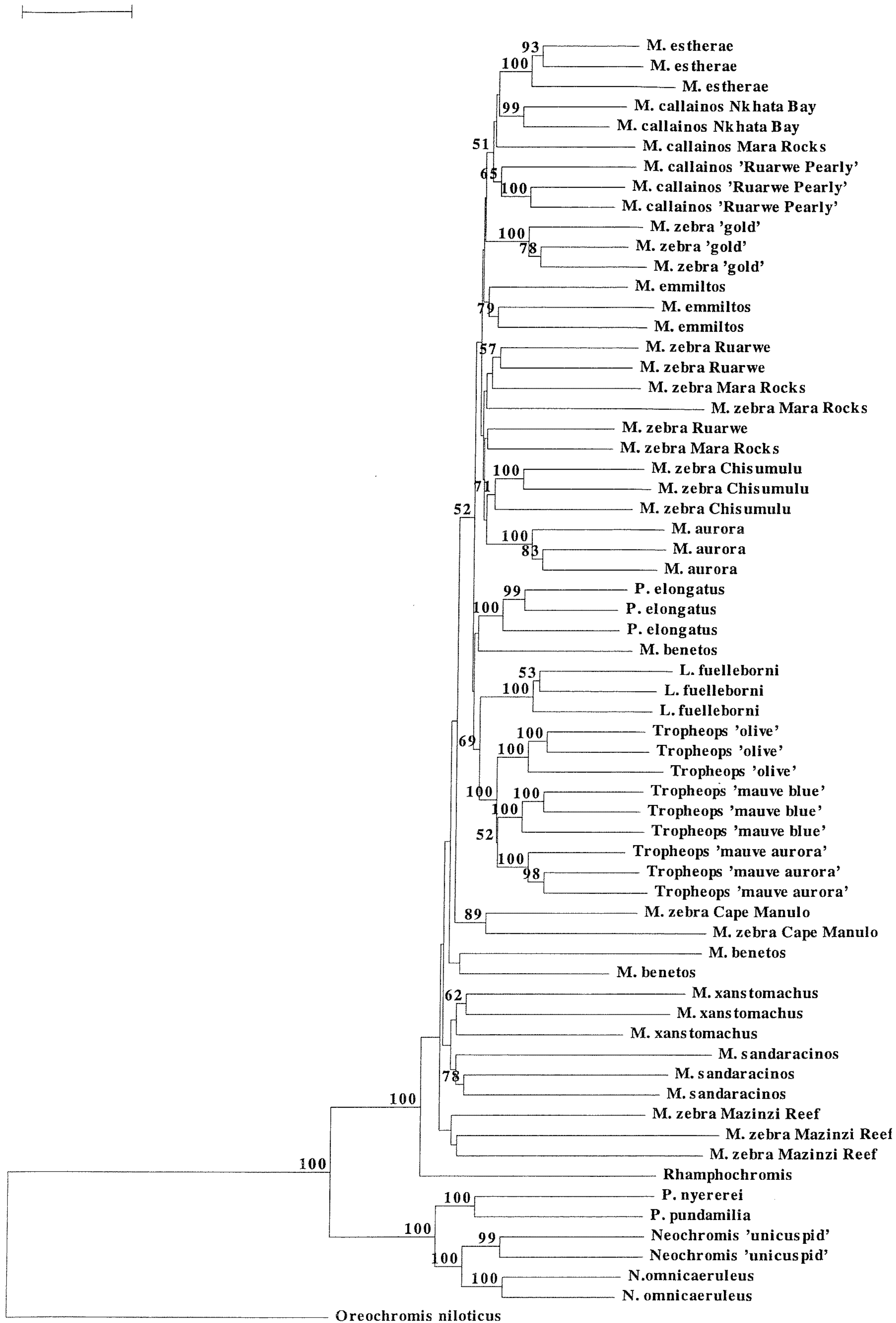


Figure 6.7. NJ tree constructed using the distance of Link *et al* (1995). Peaks were scored using Genotyper with no set peak height threshold. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support. The scale bar indicates 10% divergence in AFLP profile.

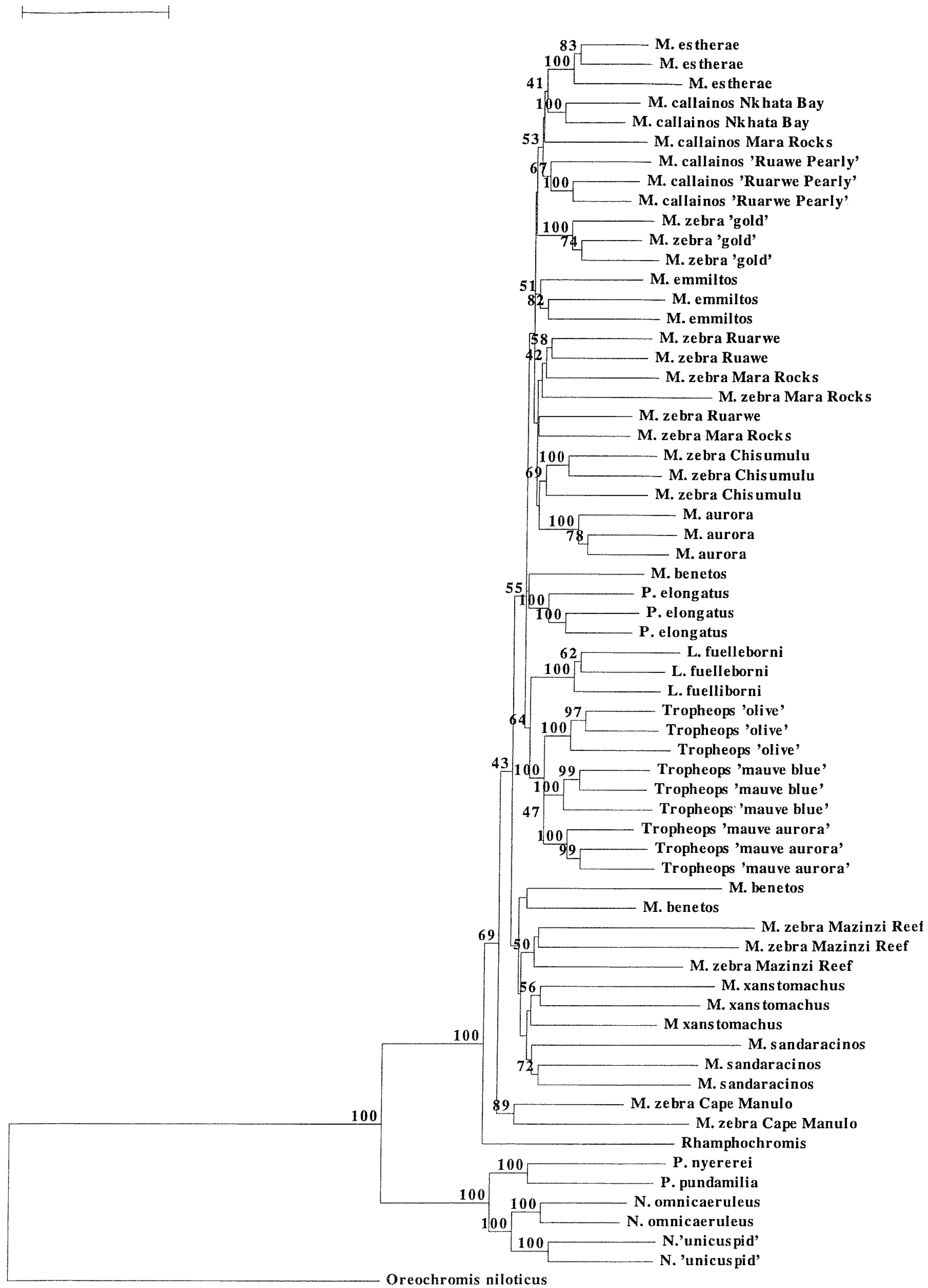


Figure 6.8. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with no set peak height threshold. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support. The scale bar indicates 10% divergence in AFLP profile.

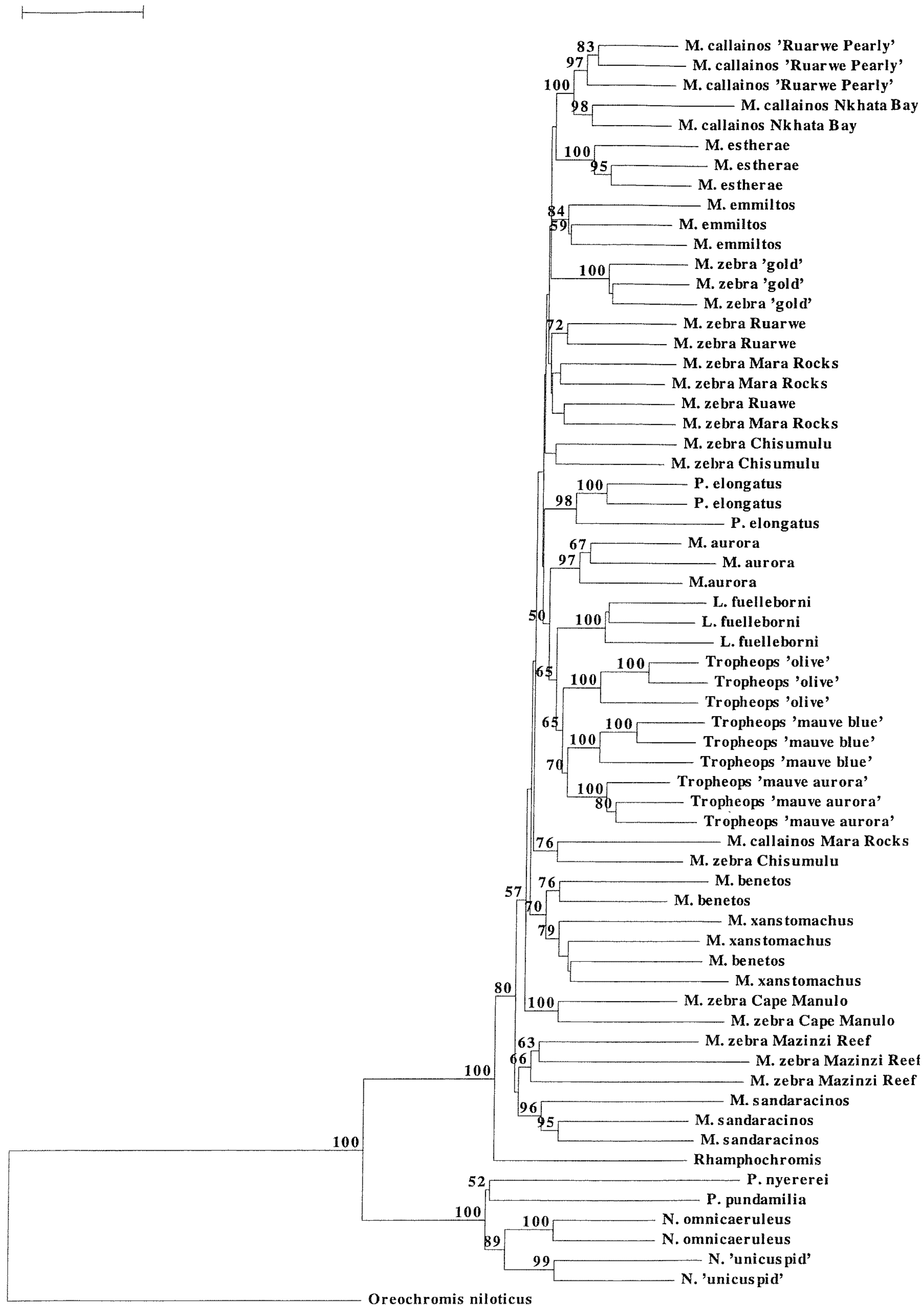


Figure 6.9. NJ tree constructed using the distance of Link *et al* (1995) Peaks were scored using Genotyper with a peak threshold height of 250 rfu. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support. The scale bar indicates 10% divergence in AFLP profile.

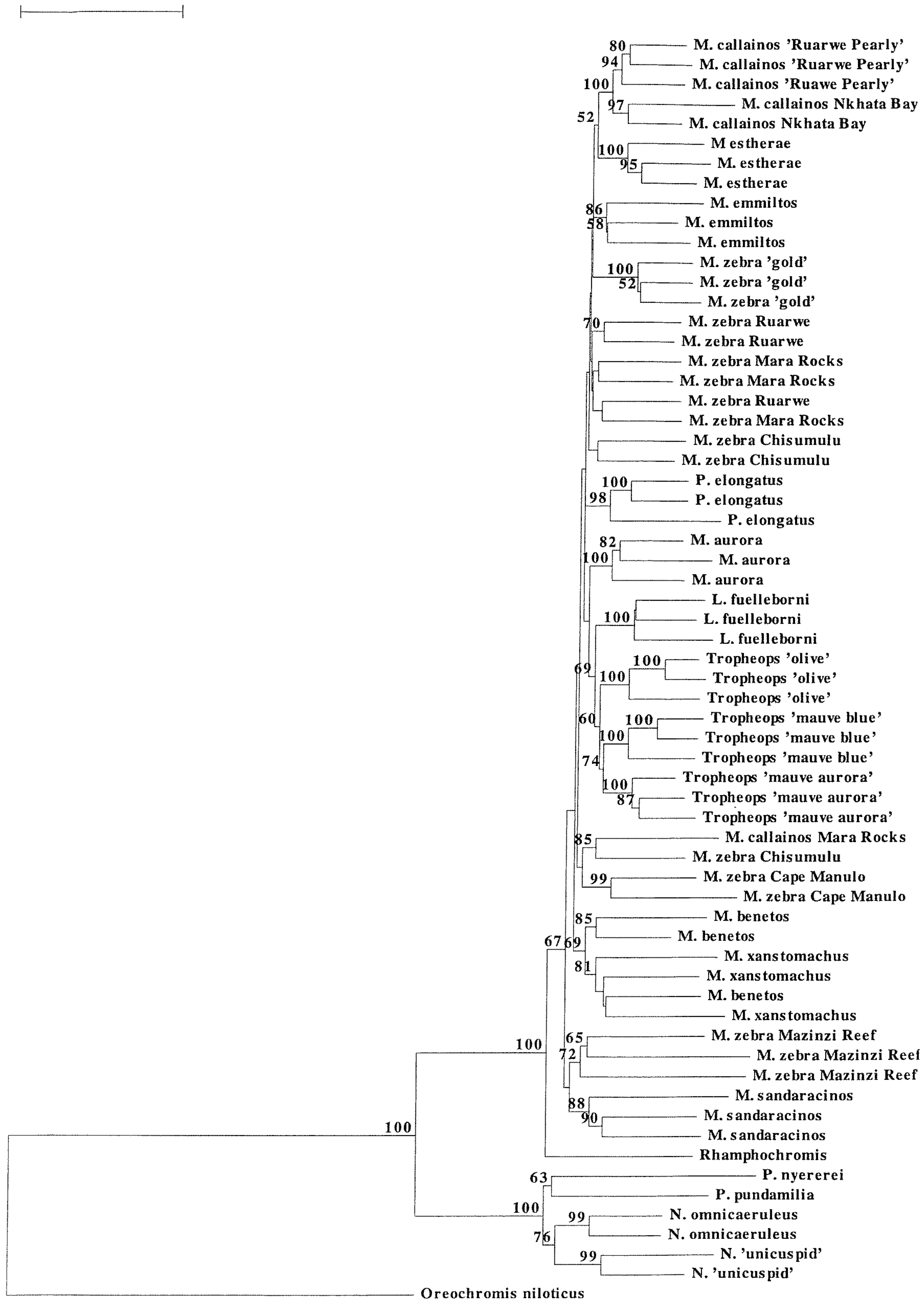


Figure 6.10. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with a peak threshold height of 250 rfu. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support). The scale bar indicates 10% divergence in AFLP profile.

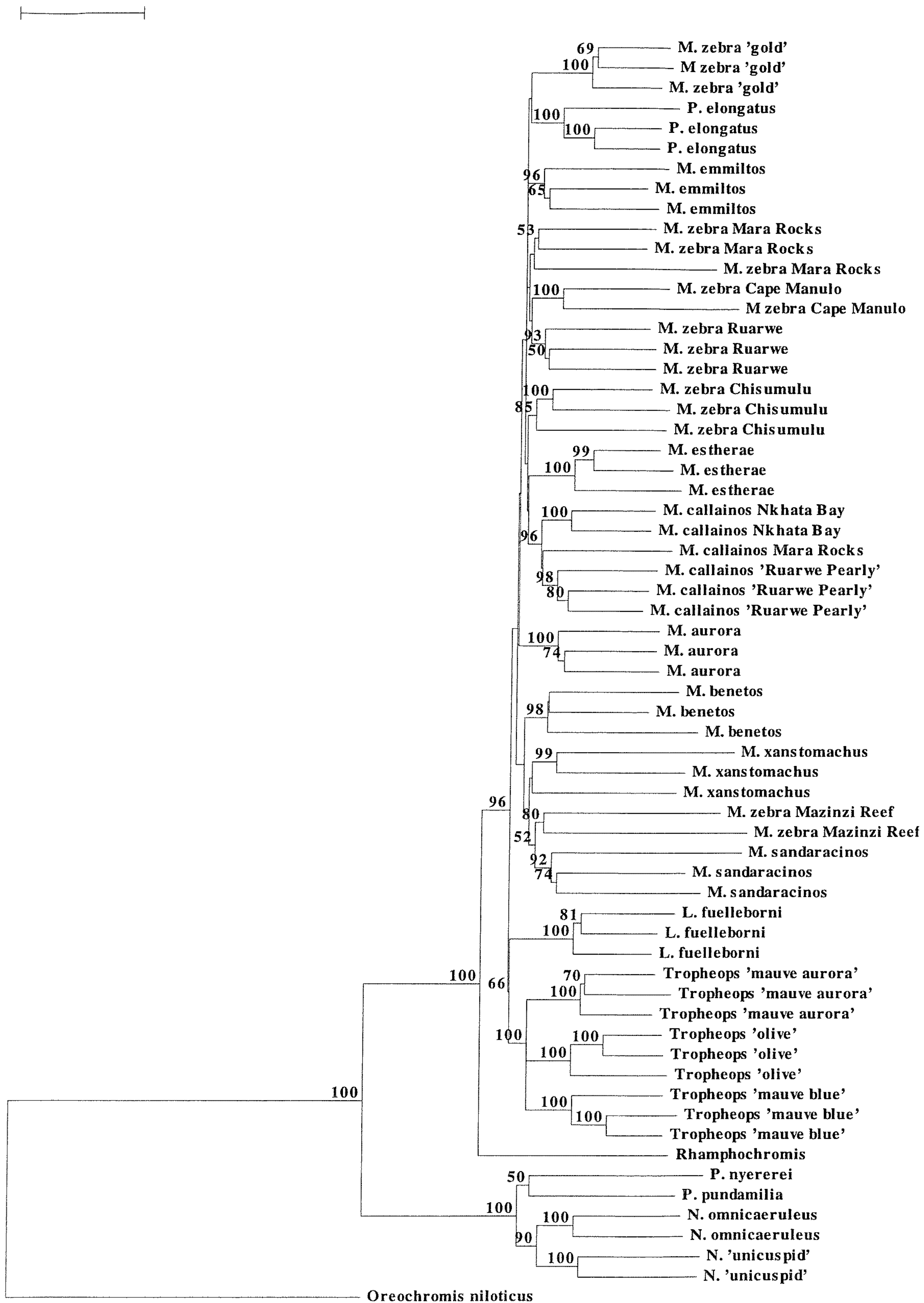


Figure 6.11. NJ tree constructed using the distance of Link *et al* (1995) Peaks were scored using Genographer. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support. Individual Zebs2 has been omitted from the analysis. The scale bar indicates 10% divergence in AFLP profile.

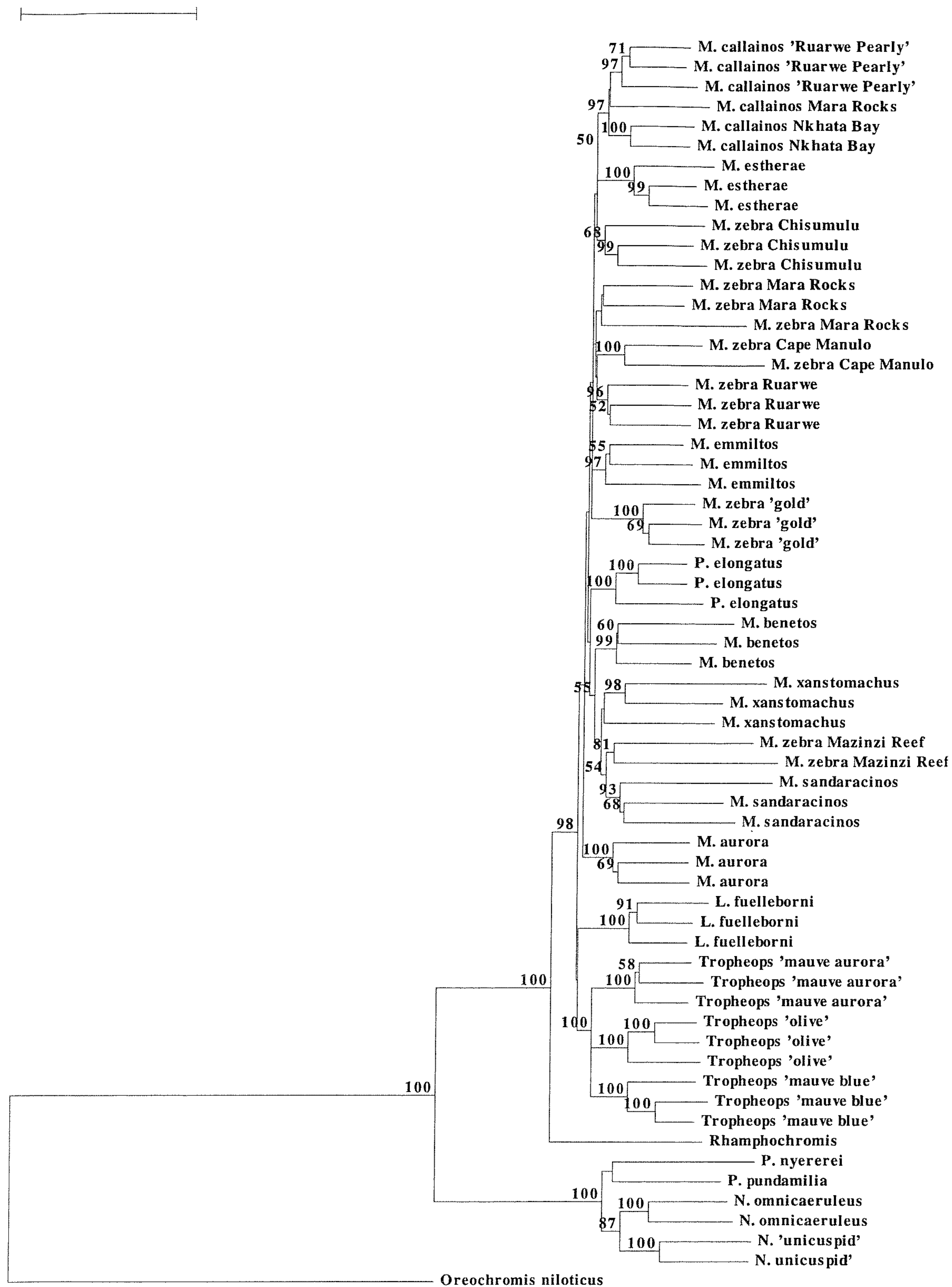


Figure 6.12. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genographer. Numbers at nodes indicate the number of times that nodes was recovered in 100 bootstrap replications (only shown for nodes with >50% support. Individual Zebs2 has been omitted from the analysis. The scale bar indicates 10% divergence in AFLP profile.

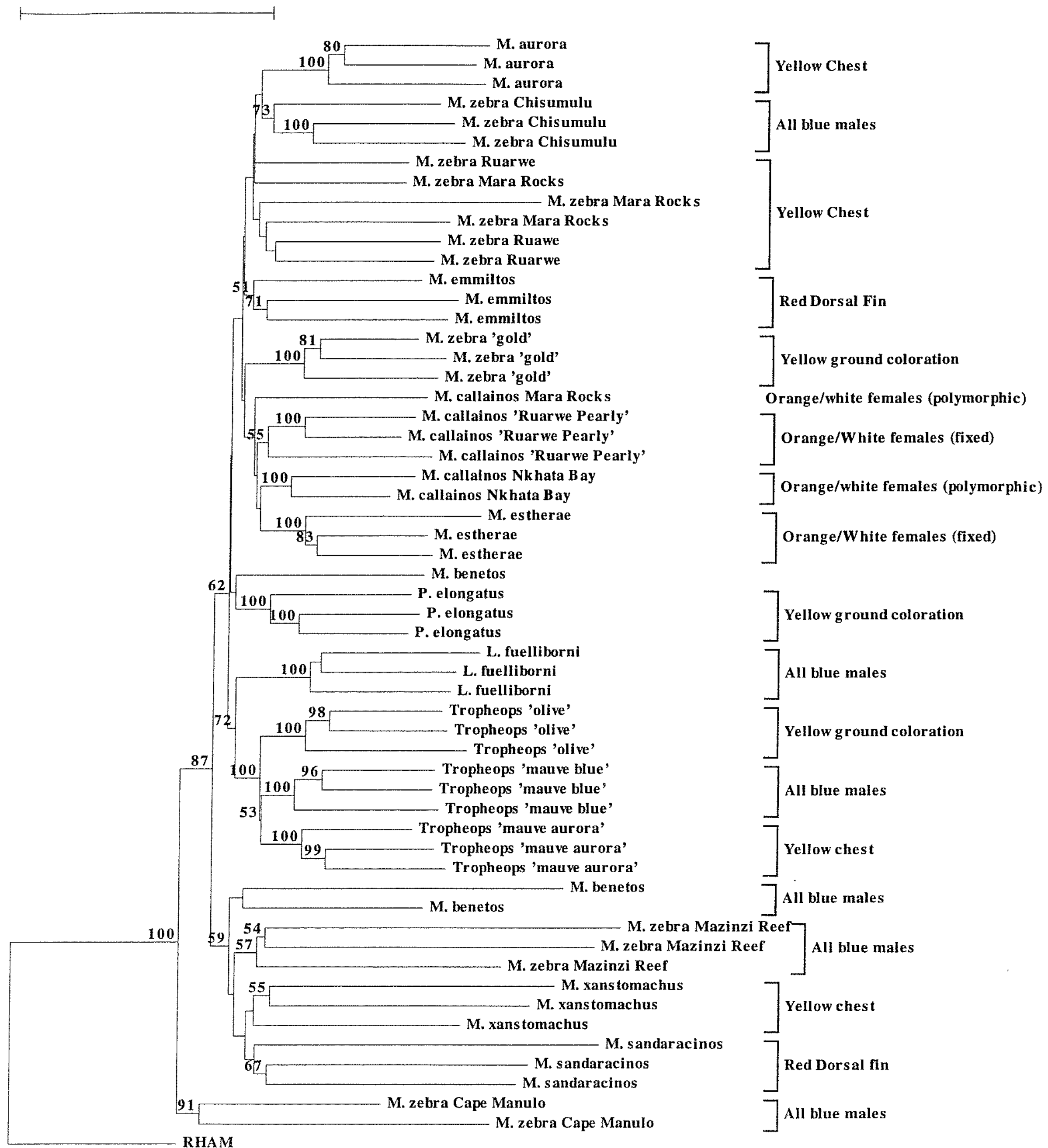


Figure 6.13. NJ tree of all mbuna populations based on peaks scored using Genotyper (manual alignment) and the genetic distance measure of Nei and Li (1979). Bootstrap values are only shown for nodes with >50% support. The colour traits of a particular population are labelled. Only populations with completely blue male coloration are labelled as such, but this trait is present in all populations except *M. zebra* 'gold', *Tropheops* 'olive', *P. elongatus*, *M. estherae* and *M. callainos* (Ruarwe). *Labeotropheus fuelleborni* is also polymorphic for white/orange females as some of the *M. zebra* populations may be.. The scale bar indicates 10% divergence in AFLP profile.

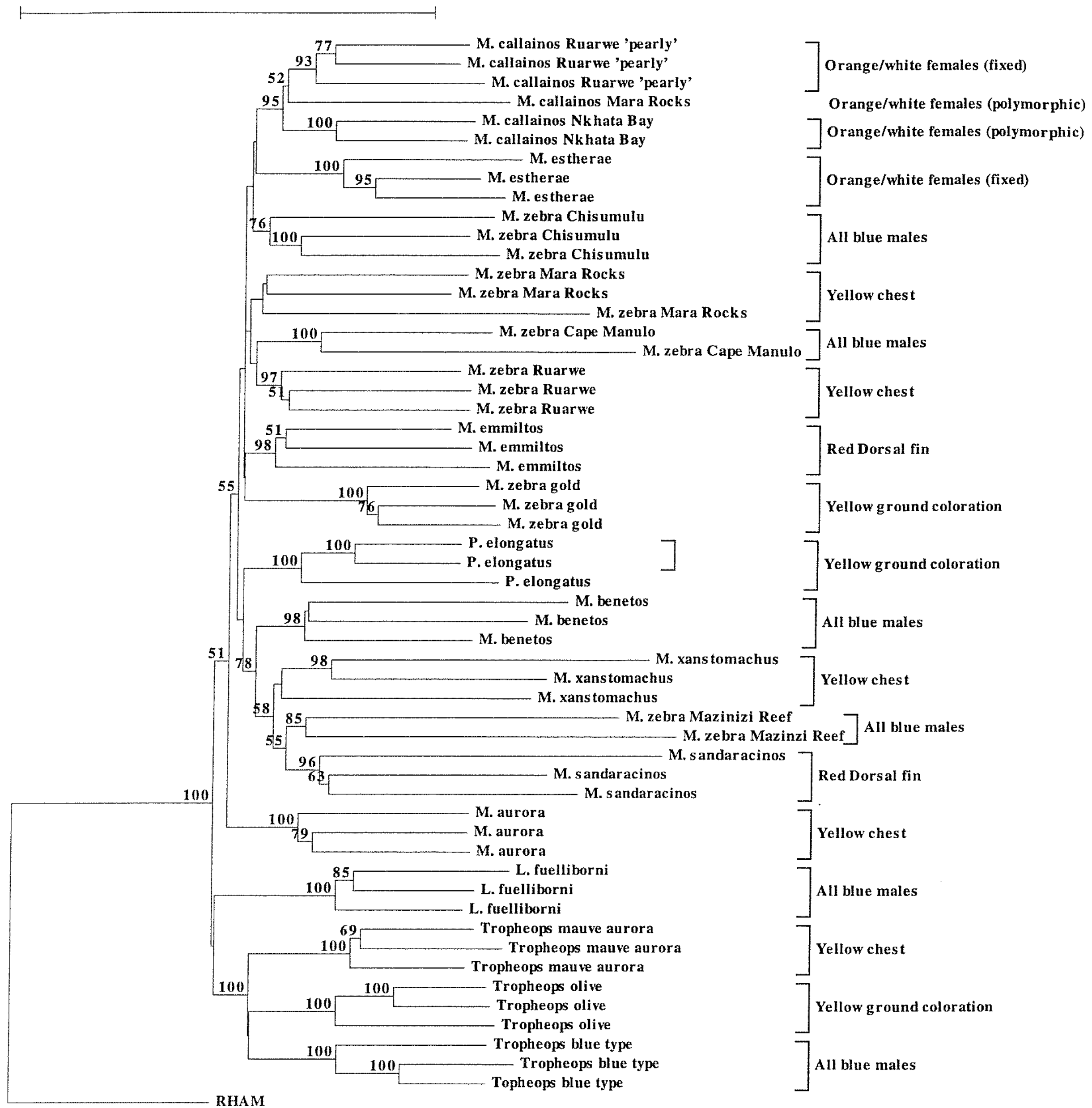


Figure 6.14. NJ tree of all mbuna populations based on peaks scored using Genographer and the genetic distance measure of Nei and Li (1979). Bootstrap values are only shown for nodes with >50% support. The colour traits of a particular population are labelled. Only populations with completely blue coloration are labelled as such, but this trait is present in all populations except *M. zebra* 'gold', *Tropheops* 'olive', *P. elongatus*, *M. estherae* and *M. callainos* (Ruarwe). Labeotropheus fuelliborni is polymorphic for white/orange females, as may be some of the *M. zebra* populations. The scale bar indicates 10% divergence in AFLP profile.

6.4 Discussion

6.4.1 Utility of AFLPs

Previously published AFLP phylogenetic studies have reported a wide variety of average numbers of bands detected per primer pair. Giannassi *et al* (2001) used 5 primer pairs with a mean of 86.8 bands per pair. A similar number of bands was detected by Parsons and Shaw (2001), who scored 70-100 bands in each of four primer pairs. Wilding *et al* (2001) scored an average of 61.2 bands from 5 primer pairs. Sharbel *et al* (2000) reported an average of just 26.3 bands from 3 primer pairs, although this figure only includes polymorphic bands. In contrast, Albertson *et al* (1999) obtained an average of 204.3 bands from each of 11 primer pairs. The mean number of bands scored per primer pair in the present study ranged from 91-141 depending on the method of peak scoring used – a relatively high range in comparison with the other studies.

Each method of peak scoring has associated advantages and disadvantages. Genographer allows manual scoring of peak presence/absence so that relatively faint bands can be detected and scored as present. It also produces a very neat gel image with no background noise allowing easy scoring. However it is a time consuming exercise – each gel takes approximately one and a half to two hours to score, depending on the number of bands generated by each primer pair. Using Genotyper to output a list of peaks for each lane and then manually aligning peaks between lanes using a spreadsheet takes a similar amount of time per gel. However, all information on the gel is used by these methods, even areas of the gel that are difficult to score manually. By far the quickest method available is to use Genotyper to automatically assign peaks to bins, as no extra manual alignment is required. A disadvantage of this method is that the number of bands scored across a gel is reduced as overlapping bins are ignored, even if the actual size range of the peaks within them do not overlap.

Variation between individuals made up a large proportion of the observed differences in AFLP profiles. This meant that it was necessary to score a large number of bands before the individuals sampled from a population clustered together, a fact reflected in Figure 6.6, which shows increasing mean bootstrap support with addition of successive

numbers of primer pairs. Relative branch lengths in all trees also are indicative of most variation occurring between individuals as inter-population or inter-clade branch lengths are all short. This pattern was also noted by Albertson *et al* (1999), and the authors suggested that short inter-clade branches are consistent with an extremely rapid radiation. The fact that both ancient (outgroups) and recent (intra-clade structure, for example in the *Tropheops* clade) as revealed by the present study suggest that AFLPs are an extremely useful class of molecular marker for phylogeny reconstruction in East African cichlids. Outgroup relationships concur with those found by other studies using mitochondrial sequence data (Meyer, 1993 and Idid *et al*, unpublished). It is possible that yet more primer pairs may provide both higher support for nodes among some mbuna populations and greater resolution as the mean bootstrap support for each node in the tree continues to increase with the addition of data from successive primer pairs (Figure 6.6). However, it is unlikely that the evolutionary relationships between all northern *Metriaclima* populations will ever be fully resolved due to the extremely rapid radiation (and hence extremely short branch lengths) of this group. Two interesting points to note are that the Victorian populations sampled are no less diverged (in terms of branch lengths) than the Malawian populations, casting doubt on the often quoted age differences between the two radiations (see Chapter 1, Table 1.1). This pattern was also observed in another AFLP based phylogenetic investigation on African cichlids (Seehausen, Koetsier and Schneider, unpublished). In addition, Owen *et al* (1990) suggested that the South East arm of Lake Malawi was dry 200 years ago, and that the endemic *M. benetos* must have evolved since then. The populations sampled from this area of the lake are no less diverged than the northern populations. This indicates that *M. benetos* and the other *Metriaclima* populations in this area must have existed prior to the low lake stand that led to the desiccation of their current locations.

6.4.2 Variations in Tree Topology

Tree topologies vary slightly with both the method of peak scoring and genetic distance measure used. Some nodes within the tree appear extremely robust – the relationships between the Victorian species, *Rhamphochromis* and the mbuna do not alter either with distance statistic or with peak scoring method. The *Tropheops* and *M. callainos* clades are also observed in all trees. However, the relationship between the *Tropheops* clade and the rest of the mbuna populations is not clear as it is shown either as a sister group

to all other mbuna populations (using Genographer) or as a sister group to just the northern *Metriaclima* and *Pseudotropheus* populations (using Genotyper). A major difference between the Genographer and Genotyper based trees is that some intra-clade relationships appear to be resolved when Genotyper is used; the trees indicate that the *M. zebra* population from Cape Manulo is basal to the rest of the mbuna populations and that the northern *Metriaclima* and *Pseudotropheus* populations are more closely related to the morphologically different *Tropheops* than to the southern *Metriaclima* populations. These relationships are supported by bootstrap values of over 50% when all populations are analysed and even higher bootstrap values when analysis is limited to *Metriaclima* populations. This tree topology is not produced when Genographer is used to score peaks; in this case the Cape Manulo *M. zebra* samples cluster with the northern *Metriaclima* populations, and the *Tropheops* clade is a sister group to all other mbuna populations. However, the southern *Metriaclima* populations still form a distinct clade. Differences in tree topologies due to distance statistics were less marked, the main difference occurs when peaks are scored using Genographer, the southern *Metriaclima* populations do not form a clade when the distance of Link *et al* (1995) is used, but do cluster together when the Nei and Li (1979) distance is employed. The phylogenetic patterns in the data are still present even when the number of bands scored is more than halved by setting a peak height threshold of 400 rfu in Genotyper (see figure 6.17 in appendix). This may indicate that very little information is being contributed by the peaks of lower intensity. However it is noted that in some of the trees produced using peak thresholds in Genotyper some individuals samples cluster with other populations. Interestingly, the trees produced by manually aligning all peaks in Genotyper are very similar in structure to those produced using peaks greater than 50, 150, or 250 rfu in height.

Regardless of which methods are used to construct the tree, it is clear that parallel evolution of colour traits has occurred in Lake Malawi cichlids. Yellow-chested populations occur in at least two different clades (*Tropheops* 'mauve aurora' in the *Tropheops* clade and *M. xanstomachus* in the southern *Metriaclima* clade) and four further populations (*M. zebra* from Ruarwe and Mara Rocks, *M. aurora* and *P. elongatus*) do not group with either of these clades, suggesting further independent evolution. That *Tropheops* populations cluster to the exclusion of others is not surprising as the *Tropheops* genus (previously described as a sub-genus by Trewavas,

1984) is a distinct group sharing particular jaw and head morphological characteristics. Likewise, yellow ground coloration has arisen more than once as it is present in the *Tropheops* clade (*Tropheops* 'olive') and also in other populations not present in this clade (*M. zebra* 'gold' and *P. elongatus*). Within the *Tropheops* clade, it is difficult to determine the relationships between the yellow chested 'mauve aurora', the all blue 'mauve' and the all yellow 'olive' as Figure 6.13 indicates 'olive' is basal to the other two but Figure 6.14 shows 'mauve aurora' occupies this position. The apparently widespread distribution of blue coloration among different clades within the mbuna suggests that this trait has evolved in parallel several times. However, an alternative explanation is that blue ground coloration may be an ancestral state rather than derived. A similar pattern of distribution among clades is also observed for the yellow chest trait, however the species/populations scored as possessing yellow chests were also mostly scored as having blue ground coloration.

Bright orange or white females occur in two clades in all trees, suggesting that this trait has evolved in parallel. As the *M. estherae* population and *M. callainos* population are both fixed for orange and white respectively, it is possible that fixation has occurred in parallel within this clade. The other *M. callainos* populations are polymorphic for female colour and have blue as well as white females. Figure 6.13 shows *M. estherae* to be paraphyletic with the rest of the *M. callainos* clade, which suggests fixation occurred twice. However, Figure 6.14 shows *M. estherae* has a sister group relationship with the other *M. callainos* populations, in this case fixation of orange and white females may have occurred once or twice. Further investigation of the origin(s) of the fixation of this trait is obviously desirable. It is possible that polymorphism for orange/white females may be an ancestral trait as it may be shared by some of the *M. zebra* populations, and thus be present in the main clades in the tree. However, fixation of this trait has only occurred in the *M. callainos* clade.

Albertson *et al* (1999) found that male coloration varied within genera but that trophic morphology did not, suggesting a secondary radiation of male colour patterns after a primary diversification in feeding mechanisms. This study also finds a similar pattern as trophic morphology does not vary within clades as each clade is composed of members of the same genus. However, there appear to be phylogenetic differences between populations currently placed in the same genus with similar trophic

morphologies. The southern *Metriaclima* populations form a well-supported clade, to the exclusion of populations with similar colour patterns. In addition, the *M. zebra* population from Mazinzi reef is present in this clade whereas the supposedly conspecific northern *M. zebra* populations are not. A previous allozyme study (McKaye *et al*, 1984) also found significant divergence between *M. zebra* populations from the north (Nkhata Bay and Chilumba, a location close to Mpanga Rocks) and south (Mumbo and Domwe Islands, situated slightly to the west of the southern locations sampled in this study). Likewise, *M. sandaracinos* is present in this clade, whereas the northern *M. emmiltos* is not. Both of these species possess red dorsal fins, indicating that this colour trait has possibly evolved more than once.

6.4.3 Why have some male colour traits evolved in parallel?

Deutsch (1997) observed that male coloration in mbuna species was biased to blue and yellow and suggested three hypotheses to account for this. Firstly, it may be due to the physical limits of the environment; clear water transmits blue light most effectively, and yellow objects contrast with reflective blues by appearing dark (Lythgoe, 1988). Secondly, cichlid physiology may restrict the number of colours that may be displayed and thirdly, blues and yellows may be dominant due to properties of the visual perception systems of females. Interestingly Carleton *et al* (2000) found that *M. zebra* is able to detect ultraviolet wavelengths. Under these conditions, the blue coloration seen in many mbuna is highly reflective (Kornfield, 1991). Deutsch (1997) found no evidence for particular male colours being associated with either depth or habitats and concluded that the variety of patterns observed were probably generated by female sexual selection.

It is possible that the two clades consisting of the *Metriaclima* populations from the north and south of the lake shown in Figures 6.13 and 6.14 are the product of current or previous localised introgression between populations. For example, the red dorsal fin trait may have evolved only once, with subsequent dispersal and population extinction to account for the current large geographic distance between the two species. A period of geneflow between the two red dorsal populations and neighbouring non red dorsals could give the impression of parallel evolution. This alternative explanation cannot be discounted with the present data set, although it seems unlikely to be true for all traits

tested in this study. As discussed previously, *Tropheops* species are morphologically and ecologically differentiated from *Metriaclima* species and it is unlikely that enough introgression could take place to cause the apparent similarities in colour patterns.

6.4.4 Parallel Speciation as well as Parallel Evolution?

The trees constructed using peaks scored with Genotyper (with manual alignment or a peak height threshold) suggest the intriguing possibility that not only colour patterns have evolved in parallel but that ecological and morphological characteristics have too. The northern *M. zebra* populations appear more closely related to the apparently divergent *Tropheops* clade than they do to the *M. zebra* population from Mazini Reef. However, not all methods of peak scoring result in a deep split within the *Metriaclima* clade and consequently further analysis of the AFLP data is required before any firm conclusions can be drawn. In addition, *Tropheops* populations from a southern location should be typed using the same primer pairs to determine if a deep split also exists in this taxon as well. A combination of morphological, behavioural, ecological and molecular data led Rundle *et al* (2000) to conclude that parallel speciation was responsible for the presence of two ecotypes (limnetic and benthic) of threespine sticklebacks in different lakes. Mate choice tests demonstrated reproductive isolation between ecotypes, but no isolation between members of the same ecotype from either the same or different lakes. Essentially the same forms arose independently in different lakes, probably due to the similar selection pressures that operate in each lake. Johannssen (2001) suggests that parallel speciation provides strong evidence for the occurrence of sympatric speciation. However the geographical mode of origin of the mbuna species is unknown and both allopatric and sympatric models have been put forward.

6.4.5 Conclusions

This study provides the first phylogenetic evidence for parallel evolution of colour patterns in Lake Malawi mbuna and indicates the usefulness of AFLPs for phylogenetic studies in this group of fish. Parallel evolution of colour traits may have occurred due to a combination of mbuna physiological and genetic limitations, and similar selection

pressures (interactions between visual signals and environmental parameters) operating in ecologically similar habitats around the lake.

6.5 References

Albertson RC, Markert JA, Danley PD and Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of. Sciences of the USA* 96 5107-5110

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ and Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelliborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London B* 266: 119-130

Carleton KL, Harosi FI and Kocher TD (2000) Visual pigments of African cichlid fishes: evidence for ultravioley vision from microspectrophotmetry and DNA sequences. *Vision Research* 40: 879-890

Deutsch JC (1997) Colour diversification in Malawi cichlids: evidence for adaptation, reinforcement or sexual selection. *Biological Journal of the Linnean Society* 62: 1-14

Estoup A and Angers B (1998) Microsatellites and minisatellites for molecular ecology : theoretical and empirical considerations. In *Advances in Molecular Ecology* ed GR Carvalho. IOS Press.

Felsenstein J (1985) Confidence limits on phylogenies, an approach using the bootstrap. *Evolution* 783-791

Futuyama DJ (1986) Evolutionary Biology. Sinauer Inc, Sunderland

Garnhardt N and Kocher TD (submitted) BinThere: a program for the analysis and automatic binning of fluorscent AFLP data.

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

Hillis DM and Huelsenbeck JP (1992) Signal, noise and reliability in molecular phylogenetic analysis. *Journal of Heredity* 83: 189-195

Johannesson K (2001) Parallel speciation: a key to sympatric divergence. *Trends in Ecology and Evolution* 16: 148-153

Knight ME, Turner GF, Rico C, van Oppen MJH and Hewitt GM (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology*: 1605-1610

Kornfield I (1991) Genetics in Keenleyside. In SMHA, *Cichlids fishes: behaviour, ecology and evolution.* Chapman and Hall, London

Link W, Dixens C, Singh M, Schwall M and Melchinger AM (1995) Genetic diversity

in European and Mediterranean faba bean germ plasm revealed by RAPD markers. *Theoretical and Applied Genetics* 90: 27-32

Lythgoe JN (1988) Light and vision in the aquatic environment. In Atema *et al* eds *The Sensory Biology of Aquatic Animals*. Springer, London

McKaye KR, Kocher T, Reinthal P, Harrison R and Kornfiled I (1984) Genetic evidence for allopatric and sympatric differentiation among colour morphs of a Lake Malawi cichlid fish. *Netherlands Journal of Zoology* 38: 215-219

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution* 8: 279-284

Nei, M.,Li, W.-H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the USA* 76: 5269-5273.

Owen BR, Crossley R, Johnson TC, Tweddle D, Kornfield I, Davison S, Eccles DH and Engstrom DE (1990) Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proceedings of the Royal Society of London B* 240: 519-553

Parsons YM and Shaw KL (2001) Species boundaries and gentic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology* 10: 1765-1772

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC and Sharp BJ (1983) A preliminary survey of the cichlid fishes of rocky habitats of Lake Malawi. *S. Afr. J. Sci.* 18: 149-310

Rico C *et al* (1992) Male reproductive tactics in the threespine stickleback – an evaluation by DNA fingerprinting. *Molecular Ecology* 1, 79-87

Rundle HD, Nagel L, Wenrick Bougham J and Scluter D (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science* 287: 306-308

Seehausen O, van Alphen JJM and Lande R (1999a) Colour polymorphism and sex ration distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecology Letters* 2: 367-378

Seehausen O, van Alphen JJM, Witte F (1999b) Can ancient colour polymorphisms explain why some cichlid lineages speciate rapidly under disruptive sexual selection? *Belgian Journal of Zoology* 129: 43-60

Seehausen O, van Alphen JJM, Witte F (1997a) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277: 1808-1811

Sharbel TM, Haubold B and Mitchell-olds T (2001) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonisation of Europe. *Molecular Ecology* 9: 2109-2118

Toonen RJ (submitted) Increased throughput for fragment analysis on an ABI 377 automated sequencer using a membrane comb and STRand software.

Van de Peer Y and de Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees in the Microsoft Windows environment. *Computer Applications in the Biosciences* 10 569-570

Vos P, Hogers R, Bleeker M (1995) AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Wilding CS, Butlin RK and Graham J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of *Evolutionary Biology* 14: 611-619

6.6 Chapter 6 Appendix

Table 6.6. Population designations used in the trees in this appendix.

Species	Collection locale	ID
Metriaclima callainos	Nkhata Bay	MCALNK
M callainos	Mara Rocks	MCALMR
M callainos 'pearly'	Ruarwe	RPEARL
M. estherae	Mozambique	ESTER
M. benetos	Mazinzi Reef	BEN
M. zebra 'gold'	Nkhata Bay	ZGOLD
Pseudotropheus elongatus	Mara Rocks	ELONG
Tropheops species'olive'	Nkhata Bay	OLIVE
M. emiltos	Mpanga Rocks	MEMIL
M. sandaracinos	Nkhudzi hills	SAND
M. aurora	?	AU
M. zebra	Chisumulu	MZCHI
M. zebra	Mara Rocks	MZMR
M. zebra	Ruarwe	MZR
M. zebra	Cape Manulo	ZEBCM
M. xanstomachus	Kanjedza Island	XAN
M. zebra	Mazinzi Reef	ZEBS
Tropheops 'mauve aurora'	Ruarwe	TROMA
Tropheops 'mauve blue'	Ruarwe	TROMB
Labeotropheus fuelleborni	Lab stock	LAB
Oreochromis niloticus	Lab stock	TILAPIA
Rhamphochromis sp	Lake Malawi	RHAM
Neochromis omnicaeruleus	Lake Victoria	NEOO
Neochromis species unicuspid scraper	Lake Victoria	NEOU
Pundamilia pundamilia	Lake Victoria	PUNDB
P. nyererei	Lake Victoria	PUNDR

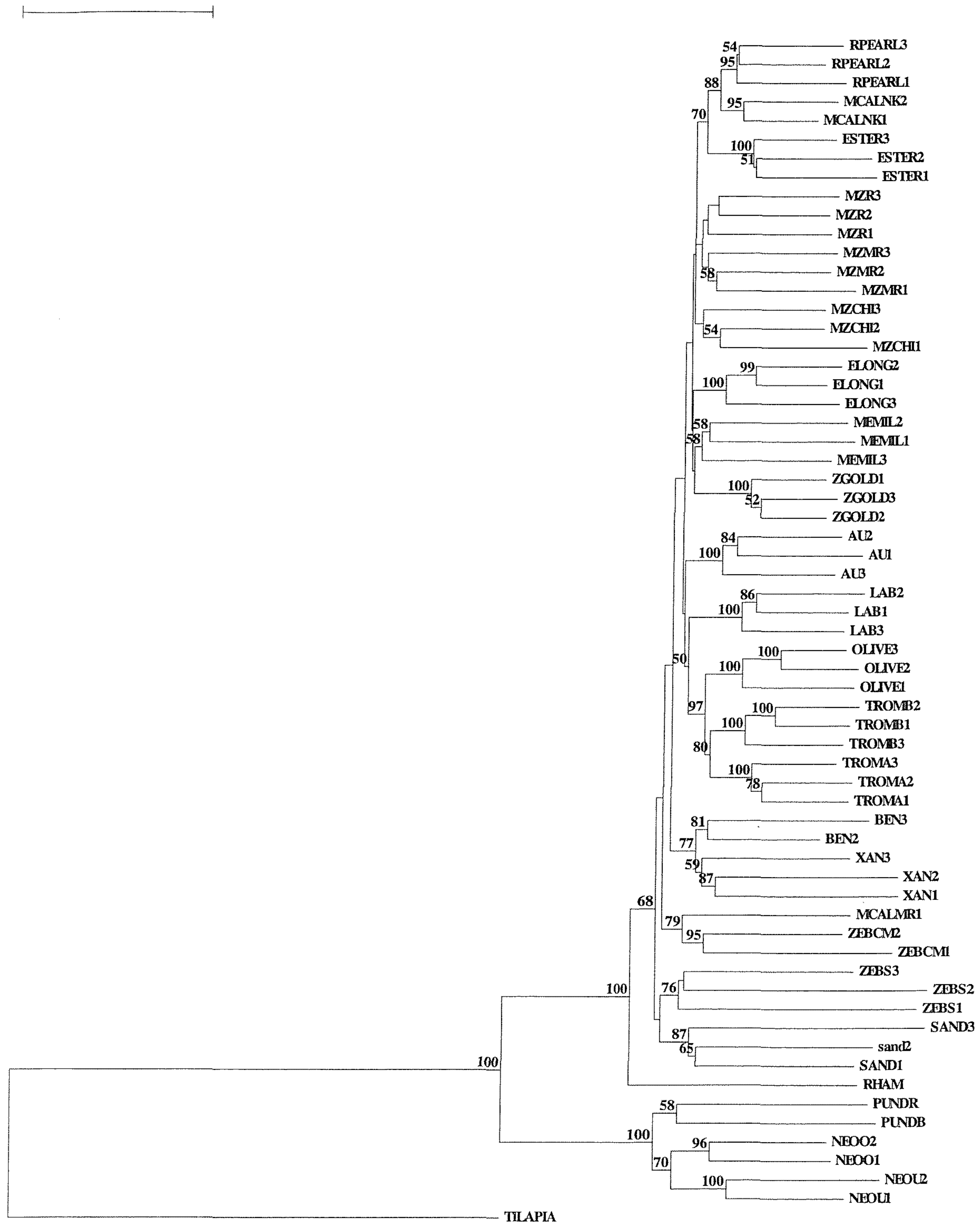


Figure 6.15. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with a peak height threshold of 50 rfu; a total of 2012 bands were scored using 18 primer pairs. Individual Ben1 has been omitted from the analysis. The scale bar indicates 10% difference in AFLP profile.

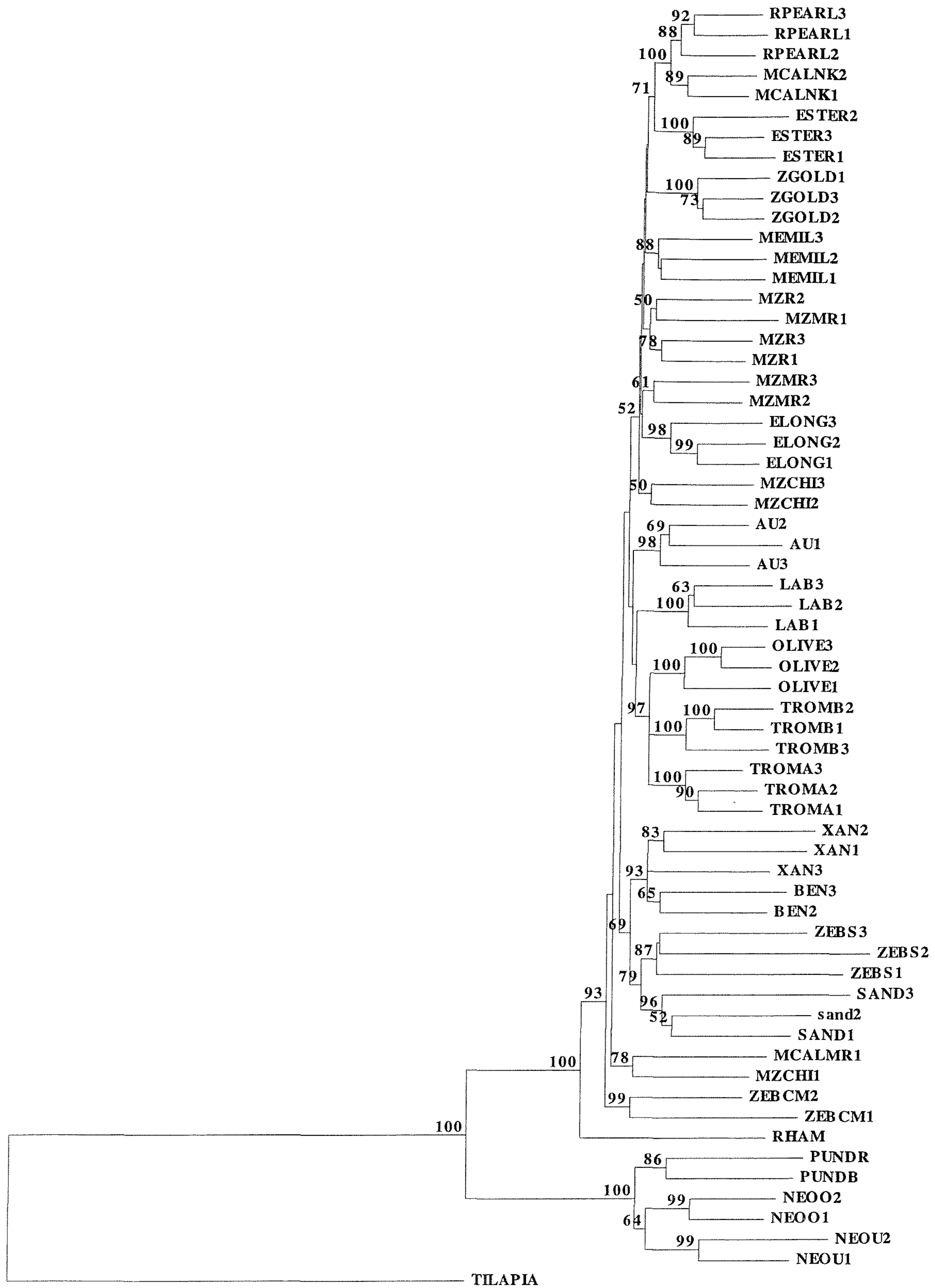


Figure 6.16. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with a peak height threshold of 150 rfu; a total of 1354 bands were scored using 18 primer pairs. Individual Ben1 has been omitted from the analysis. The scale bar indicates 10% difference in AFLP profile.

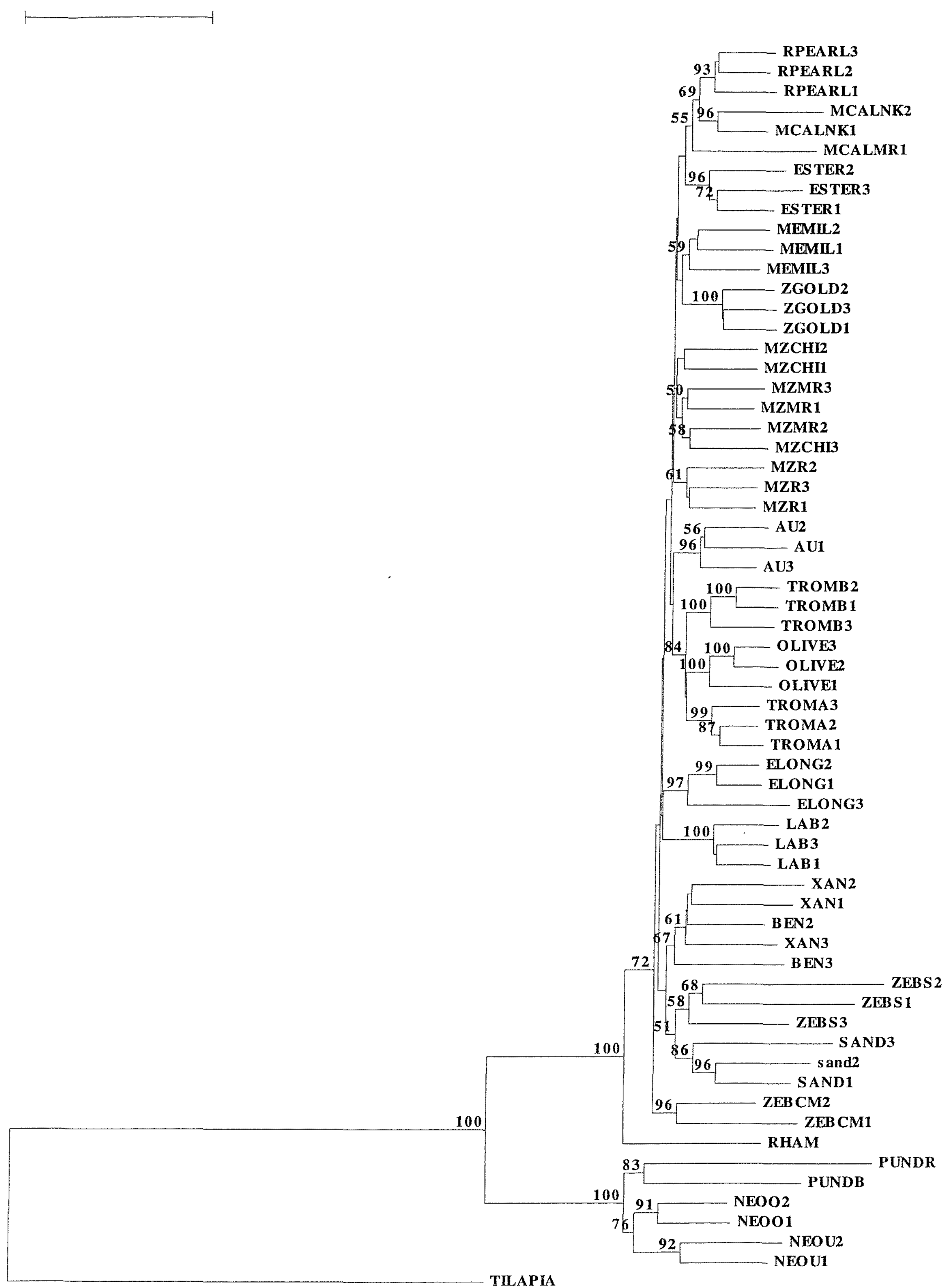


Figure 6.17. Nei 400. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with a peak height threshold of 400 rfu; a total of 950 bands were scored using 18 primer pairs. Individual Ben1 has been omitted from the analysis. The scale bar indicates 10% difference in AFLP profile.

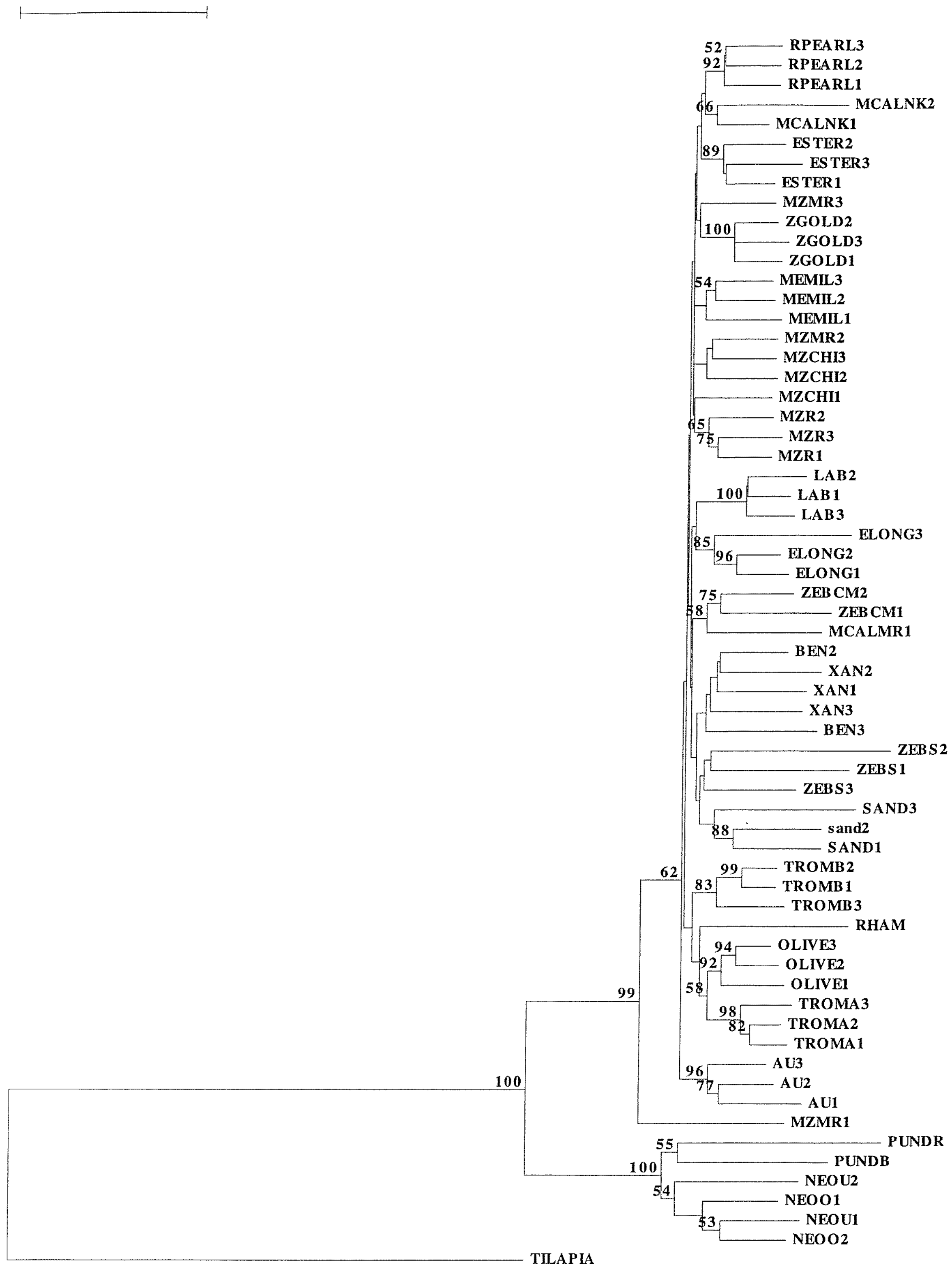


Figure 6.18. NJ tree constructed using the distance of Nei and Li (1979). Peaks were scored using Genotyper with a peak height threshold of 600 rfu; a total of 610 bands were scored using 16 primer pairs (there were no bands of this height produced by primer pairs 10 and 11). Individual Ben1 has been omitted from the analysis. The scale bar indicates 10% difference in AFLP profile

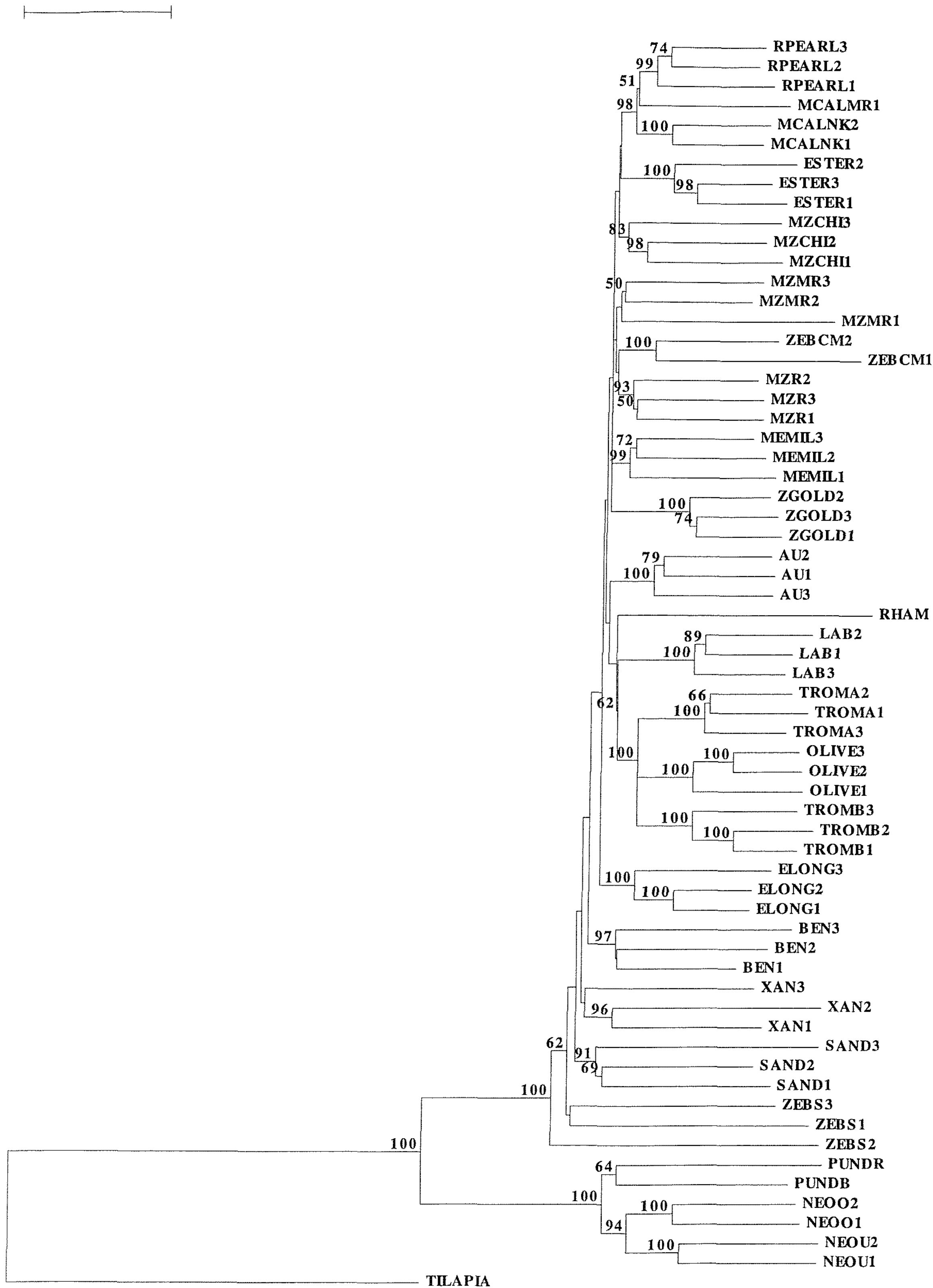


Figure 6.19. NJ tree based on data scored using Genographer and the distance measure of Link *et al* (1995). The figures at each node represent the number of times that node was recovered in trees constructed from 100 bootstrap resamplings of the data. Only nodes with greater than 50% support are labelled. The scale bar indicates 10% divergence in AFLP profile. If included, sample ZEBS2 alters the toplogy of the tree and displaces *Rhamphochromis* as the outgroup to the mbuna.

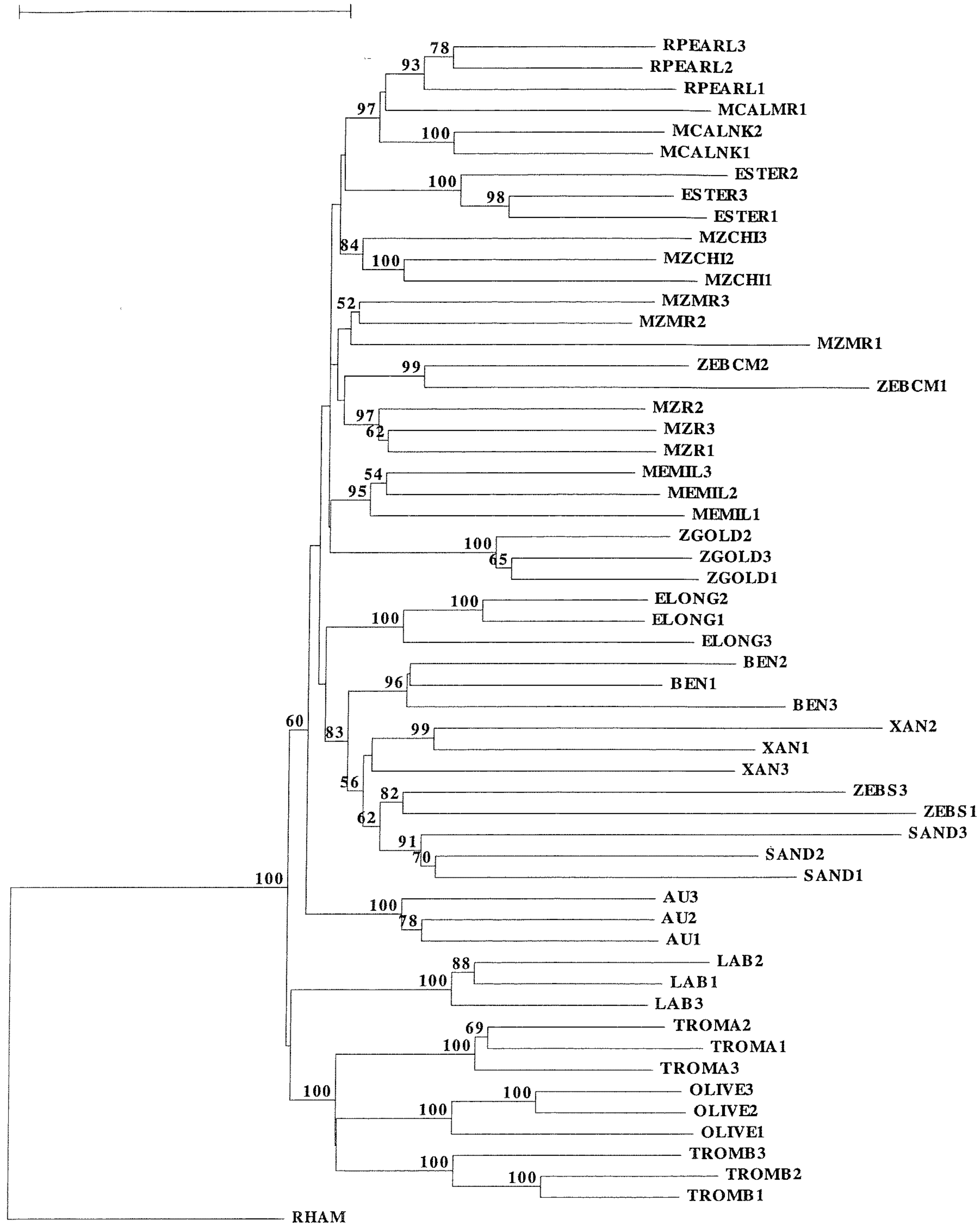


Figure 6.20. NJ tree based on data scored using Genographer and the distance measure of Link *et al* (1995). The figures at each node represent the number of times that node was recovered in trees constructed from 100 bootstrap resamplings of the data. Only nodes with greater than 50% support are labelled. Individual Zebs2 has been omitted. The scale bar indicates 10% divergence in AFLP profile.

7 General Discussion

This thesis has contributed useful information to the body of knowledge regarding the evolution of African cichlid fish. In addition, the research described has indicated several interesting avenues worthy of further investigation, which will be described in further detail below.

Laboratory tests of mate choice have already proved to be useful tools in a number of situations, from demonstrating reproductive isolation between sympatric colour morphs (Knight *et al*, 1998), determining the genetic basis of female preference and male colour traits (Seehausen *et al*, 1999) and demonstrating that colour is an important cue for females when discriminating among potential mates (Seehausen and van Alphen, 1998). Mate choice trials can also be used to ascertain the specific status of allopatric populations that exhibit variation in male breeding colours, as described in Chapter 3. Although inconclusive, the results of this study indicate that non-assortative mating occurred between *M. zebra* from Ruarwe and Nkhata Bay. Taken at face value, this indicates that the current conspecific status of these two populations is correct. However, due to technical problems with the PCR amplification of DNA from eggs it was not possible to assess whether male-male interactions (behavioural dominance) were a problem in these experiments. Male-male interactions were documented during trials involving a sympatric species pair from Lake Victoria (Chapter 4), and these interactions appeared to influence female choice. The results described in Chapter 4 also hint at the fragile nature of prezygotic reproductive barriers between closely related species of African cichlid. Assortative mating may be disrupted when a heterospecific male is larger or more active than the conspecific male present. These results indicate that laboratory mate choice trials should be carefully designed and preferably present females with more than one representative of each species to choose among. A large partitioned experimental tank will help prevent problems of male dominance (as discussed in Turner *et al*, 2001), as each male may maintain an adequate territory. Chapter 4 also indicated that microsatellite typing of partial broods and behavioural indices are both good indicators of mate preference, and both indicated similar patterns of preference among *Pundamilia* F1 hybrid females. F1 hybrid female preference was revealed to lie between the preferences of females of the parental

species, suggesting that preference for 'blue' males (*P. pundamilia*) is incompletely dominant over preference for 'red' (*P. nyererei*). Such a genetic basis effectively rules out the model of speciation described by Turner and Burrows (1995) for this particular species pair as under the model, speciation occurs when mate preference is under the control of a single, dominant gene.

An empirical investigation of the effect of microsatellite locus number and variability on a statistical estimate of relatedness revealed that using higher numbers of more variable loci produced accurate estimates with a lower variance (Chapter 2). This concurs with the results of previously published studies (Blouin *et al*, 1996; Lynch and Ritland,1999). However, simulation and rarefaction analyses revealed that nine loci provided estimates of a similar accuracy to those obtained from the eleven loci available. This has important implications for future studies of relatedness in any species as determining the minimum number of loci required in advance could have a noticeable impact on the time and money required to complete such investigations. If population allele frequency data is already available for the species of interest then simple simulations can be used to compare the performance of various groups of loci.

The rapidity of the evolution of the mbuna group of Lake Malawi has thwarted many previous attempts at phylogenetic resolution (e.g. Moran and Kornfield, 1993). As pointed out by Kornfield and Smith, 2000), the lack of a robust phylogeny has hindered evolutionary studies of this group. Chapters 5 and 6 report the phylogenetic potential of two classes of molecular markers, microsatellites and AFLPs. The microsatellite phylogenies described in Chapter 5 are poorly resolved, probably due to the relatively small number of loci used. However a number of striking similarities may be observed between the microsatellite and the AFLP phylogenies. Of all genera studied, *Tropheops* appears to be the most genetically distinct as conspecific populations of *Tropheops* species 'mauve' and 'olive' clustered together, often to the exclusion of others. All *Tropheops* species form a distinct, well-resolved clade in the AFLP based phylogenies. *M. callainos* populations also form distinct clusters in both sets of phylogenies. The inclusion of *M. estherae* in this clade in the analysis of AFLP data suggests that a lack of a female morph with vertical bars has only arisen once among the populations sampled. Likewise, the distinct ecological and morphological characters associated with the *Tropheops* species appear to have only arisen once with subsequent dispersal

and diversification in colour patterns. However, the divergence between northern and southern *Metriaclima* populations means that a single origin for *Tropheops* should be tested by typing members of this genus from the southern area of the lake. The relatively distant relationship of the northern and southern *Metriaclima* populations revealed by some analyses of AFLP data suggests that similar ecolological and morphological characters may have arisen in parallel in the two regions (parallel speciation). This requires further investigation, both in terms of phylogenetic analysis, and in terms of assessing reproductive isolation between similarly coloured populations using laboratory tests of mate choice. The division between *Metriaclima* species has not been detected in other phylogenetic studies (Albertson *et al*, 1999; Chapter 5) as in these cases populations were not sampled on a wide enough geographic scale. Whether species have evolved in parallel or not, it is clear from the AFLP phylogenies that male colour patterns have arisen more than once in different lineages. This supports the conclusions of Deutsch (1997) and McElroy *et al*, 1991) who observed that colour patterns appear to be phylogenetically unconstrained. Branch lengths among the northern *Metriaclima* populations are very short, indicating populations/species diverged over an extremely short period of evolutionary time. The relationships between these species are not resolved, even with data from 18 primer pairs. It therefore unlikely that the addition of data from extra primer pairs would provide extra resolution, although further data may be of use in clustering members of the same population. While both the AFLP and microsatellite phylogenies have revealed evolutionary relationships among some mbuna species, it is difficult to infer much about the geographical mode of speciation. This is due to a lack of resolution in the microsatellite based phylogenies, and the fact that not enough heterospecific populations were sampled from the same locations for the AFLP study.

In summary, this thesis has highlighted several area of research where molecular markers have proved invaluable to the study of speciation in African cichlid fish. However, as noted by Snoeks (2000) on a review of the systematics of the Great Lake cichlids, information regarding the general biology, ecology and morphology of cichlids should not be overlooked when drawing conclusions from molecular data. This is particularly highlighted by the need for well-designed laboratory based tests of mate choice. A concerted combination of efforts across all fields of study should continue to

reveal the mechanisms responsible for generating so much diversity within this fascinating group of organisms.

7.1 Future Work

The research described in this thesis has opened up several avenues worthy of further investigation. The addition of five extra loci to those available for use in studies of the mbuna is of obvious importance. These extra loci also indicate that other loci originally isolated from other species of African cichlids may be of use in cichlids from both Malawi and Victoria. However, care must be taken that loci with null alleles (as found in Chapter 4) are not included in further research, particularly in paternity assignment studies. Non-amplifying alleles will always allow for ambiguities in the results of such studies. The extra five loci have allowed for increased reliability of relatedness estimates (Chapter 2). Relatedness estimates could be used to investigate sex biased dispersal (as in Knight *et al*, 1999) in species in which this phenomenon has not yet been demonstrated. Another area of investigation might be to use relatedness estimates to infer the heritability of male traits, for example colour pattern, in natural populations (Ritland, 2000). This may provide an alternative to laboratory breeding programs designed with the same purpose in mind.

Some of the newly optimised loci proved useful for paternity assignment experiments as their highly polymorphic nature allowed discrimination between individual males. Well-designed laboratory tests of mate choice are an important tool in determining the existence of reproductive isolation between both allopatric and sympatric populations (Turner *et al*, 2001). Further tests on female discrimination between males from allopatric populations are already under way (M.E. Knight, unpublished data). The mate choice of inter-specific hybrids is also an important topic of investigation. Mate choice tests involving different generations of hybrid females should also help to reveal the genetic basis of mate preference. Broods from a number of F2 female hybrids between *P. pundamilia* and *P. nyererei* await complete or partial typing. Assessing the mate preference of these females will provide further information on the genetic basis for mate preference in these species. Work is currently under way on an mbuna species pair and hybrids between them (Barson *et al*, unpublished data). As discussed in

Chapter 4, it may be best to avoid locus UNH130 and to use loci isolated from *Pundamilia* or closely related species to reduce the occurrence of null alleles.

Chapter 5 has indicated the level of phylogenetic information present in a six locus microsatellite data set. As discussed, future studies would obviously require a higher number of loci, and should probably be limited to members of a particular genus/species complex. The best resolved mbuna groups comprised members of the *Tropheops* genus, and these may prove to be useful species for further phylogenetic investigations. A more promising class of molecular marker for phylogenetic investigations appears to be AFLPs. These have provided resolution of mbuna species and genera. However, the trees produced in Chapter 6 are obviously not completely resolved. This may be addressed by the use of further primer pairs, however it seems likely that phylogenetic relationships among the northern *Metriaclima* populations may never be fully resolved. This may be due in part to the apparently large amount of variation within as compared to between populations. However, as demonstrated by Chapter 6, AFLPs allow hypotheses regarding patterns of evolutionary change to be rigorously tested, which has not been possible until now. Further studies using AFLPs should perhaps consider using a larger samples size to enable hierarchical comparisons of genetic diversity such as Analysis of Molecular Variance (AMOVA). Recent developments have also indicated that loci affected by selection can skew phylogenetic analysis; Wilding *et al* (2001) used computer simulations of fixation indices to identify AFLP loci under high selection pressure. This represents a fascinating avenue of investigation and work is currently under way to determine if the same analytical methods can be applied to the data from Chapter 6.

7.2 REFERENCES

Albertson RC, Markert JA, Danley PD and Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of. Sciences of the USA* 96 5107-5110

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

Deutsch JC (1997) Colour diversification in Malawi cichlids: evidence for adaptation, reinforcement or sexual selection. *Biological Journal of the Linnean Society* 62: 1-14

Knight ME, Turner GF, Rico C, van Oppen MJH and Hewitt GM (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology*: 1605-1610

Knight ME, van Oppen MJH, Smith HL, Rico C, Hewitt GM and Turner GF (1999) Evidence for male biased dispersal in Lake Malawi cichlids from microsatellites. *Molecular Ecology* 8: 1521-1527

Kornfield I and Smith PF (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Review of Ecology and Systematics* 31: 163-196

Lynch M and Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753-1766

McElroy DM, Kornfield I and Everett J (1991) Coloration in African cichlids: diversity and constraints in Lake Malawi endemics. *Netherlands Journal of Zoology* 41: 250-268

Moran P and Kornfield I (1993) Retention of an ancestral polymorphism in the mbuna species flock of Lake Malawi. *Molecular Biology and Evolution* 10: 1015-1029

Ritland K (2000) Marker inferred relatedness as a tool for detecting heritability in nature. *Molecular Ecology* 9, 1195-1204

Seehausen O, van Alphen JJM and Lande R (1999) Colour polymorphism and sex ration distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecology Letters* 2: 367-378

Seehausen O and van Alphen JJM (1998) The effect of male colouration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioural Ecology and Sociobiology* 42 : 1-8

Snoeks J (2000) How well known is the ichthyodiversity of the large East African Lakes? *Advances in Ecological Research* 31:15-37

Turner GF, Seehausen O, Knight ME, Allender CJ and Robinson RL (2001) How many species of cichlid fishes are there in African lakes? *Molecular Ecology*

Wilding CS, Butlin RK and Graham J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* 14: 611-619